One-step Local M-estimator for Integrated Jump-Diffusion Models *

Yuping Song
School of Finance and Business, Shanghai Normal University,
Shanghai, 200234, P.R.C.

Hanchao Wang†
Zhongtai Securities Institute for Financial Studies, Shandong University,
Jinan, 250100, P.R.C.

Abstract

In this paper, robust nonparametric estimators, instead of local linear estimators, are adapted for infinitesimal coefficients associated with integrated jump-diffusion models to avoid the impact of outliers on accuracy. Furthermore, consider the complexity of iteration of the solution for local M-estimator, we propose the one-step local M-estimators to release the computation burden. Under appropriate regularity conditions, we prove that one-step local M-estimators and the fully iterative M-estimators have the same performance in consistency and asymptotic normality . Through simulation, our method present advantages in bias reduction, robustness and reducing computation cost. In addition, the estimators are illustrated empirically through stock index under different sampling frequency.

MSC 2010 subject classifications: 62M05, 62G20.
Keywords: One-step local M-estimation, integrated jump-diffusion model, infinitesimal coefficients, robust estimation, financial high-frequency data.

1 Introduction

The diffusion model defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \tag{1}$$

*This research work is supported by the National Natural Science Foundation of China (No. 11371317, 11171303), the Fundamental Research Fund of Shandong University (No. 2016GN019) and the General Research Fund of Shanghai Normal University (No. SK201720).

†Corresponding author, email: wanghanchao@sdu.edu.cn

is a widely used continuous-time stochastic model in economic and finance market to depict the dynamics of an underlying assets such as log stock indices, free foreign exchange rates and commodity prices, where $\mu(x)$ and $\sigma^2(x)$ economically represents the risk-neutral return and volatility of X_t respectively. So for practical applications, estimating the coefficients $\mu(x)$ and $\sigma^2(x)$ is necessary. In fact, by Itô formula, the infinitesimal conditional moment equtations

$$\lim_{\Delta\to 0} E\left[\frac{X_{t+\Delta}-X_t}{\Delta}|X_t=x\right]=\mu(x) \tag{2}$$

and

$$\lim_{\Delta\to 0} E\left[\frac{(X_{t+\Delta}-X_t)^2}{\Delta}|X_t=x\right]=\sigma^2(x) \tag{3}$$

provide the basis for statistical inference. Given the discrete-time observations $\{X_{i\Delta_n}; i=1,2,\cdots\}$, many statisticians and economists estimated $\mu(x)$ and $\sigma^2(x)$ by using the nonparametric regression method based on (2) and (3), where Δ_n is the sampling interval tending to 0. One can refer to Bandi and Phillips [3] and Fan and Zhang [9] et al. for more details.

As we know, many stochastic processes in empirical finance or physics can be seen as integrated stochastic processes, since the observation behaves as the accumulation of all past perturbations, such as stock prices or nominal exchange rates (Nicolau [20]) and the velocity of the particle on the surface of a liquid (Rogers and Williams [22]). Especially, in some financial cases, returns are generally stationary in mean and weak in autocorrelation, and the distribution is not normal. In order to specify these phenomena, Nicolau [20] introduced the integrated diffusion model motivated by unit root processes under the discrete framework of Park and Phillips [21]

$$\begin{cases} dY_t = X_t dt, \\ dX_t = \mu(X_t)dt + \sigma(X_t)dW_t. \end{cases} \tag{4}$$

The model (4) can accommodate nonstationarity and be made stationary by differencing, which can not be performed through univariate diffusion model due to the nondifferentiability of a Brownian motion. Nicolau [20] studied the asymptotic properties of Nadaraya-Watson estimators of $\mu(x)$ and $\sigma^2(x)$. Wang and Lin [26], Wang, Zhang and Tang [28] improved the Nadaraya-Waston estimators for $\mu(x)$ and $\sigma^2(x)$ in model (4).

Recently, growing evidence in finance shows that diffusion with jumps are becoming an increasing important issue, which can accommodate the impact of sudden and large shocks to financial markets (see Johannes [17], Aït-Sahalia and Jacod [1], Bandi and Nguyen [2]). It is reasonable to extend the model (4) to the following case:

$$\begin{cases} dY_t = X_t dt, \\ dX_t = \mu(X_{t-})dt + \sigma(X_{t-})dW_t + \int_{\mathscr{E}} c(X_{t-},z)r(\omega,dt,dz), \end{cases} \tag{5}$$

where $\mathscr{E}=\mathbb{R}\setminus\{0\}$, $W=\{W_t\}_{t\geq 0}$ is standard Brownian motion, $r(\omega,dt,dz)=(p-q)(dt,dz)$, $p(dt,dz)$ is a time-homogeneous Poisson random measure on

$\mathbb{R}_+ \times \mathbb{R}$, which is independent of W_t, and $q(dt, dz)$ is its intensity measure, that is, $E[p(dt, dz)] = q(dt, dz) = f(z)dzdt$, $f(z)$ is a Lévy density. In empirical analysis part, we will use real financial data to testify the existence of jumps and prove the validity of model (5) other than (4) in helping us modeling integrated econometric phenomena.

In Song and Lin [24], Song, Lin and Wang [25], the nonparametric estimation and inference for the unknown coefficients in model (5) are discussed based on Nadaraya-Watson method and the discrete time observations $\{Y_{i\Delta_n}; i = 1, 2, \cdots\}$, where Δ_n is the sampling interval. Chen and Zhang [7] considered the local linear estimators to correct boundary bias. However, these previous estimation procedures for model (5) are obtained by using the least-squares techniques which are sensitive to outliers, and may result in larger bias and unreasonable conclusions. Meanwhile, outliers are common in economic, financial industry, and other applications such as heavy tail data, so how to deal with these observations is a very important issue in financial time series analysis. A primary purpose of the present paper is to establish a robust nonparametric statistical inference for model (5) which generates heavy tail data.

M-type regression estimators are natural candidates for achieving desirable robust properties. So far, the nonparametric M-type estimators have been already studied by many authors, such as Huber [13], Hall and Jones [12] and so on. In addition, an inordinate amount of attention has been focused on local M-estimator, which not only inherits the robustness of M-estimator, but also can achieve the bias reduction like local linear estimator does. Fan and Jiang [10] established the consistency and asymptotic normality for the local M-estimators and one-step iterative local M-estimation of the regression function and its derivative based on the samples with independent identical distribution. Jiang and Mack [16] studied them for stationary dependent case. Wang, Zhang and Tang [27] only developed local M-estimation for jump-diffusion, Wang and Tang [29] considered local M-estimators for the unknown infinitesimal moments of continuous second-order diffusion model.

The local M-estimators require iteration for numerical implementation, while the local one-step M-estimator not only retains the robustness and the same asymptotic performance of local M-estimators based on good initial value but also bring about computational expediency. In this paper, we will apply the one-step local M-estimator in Fan and Jiang [10] to the unknown coefficients in integrated jump-diffusion model (5). To our knowledge, it is the first article regarding the one-step local M-estimator for diffusion model with jump, especially integrated jump-diffusion model (5).

The paper is organized as follows. In Section 2, we introduce the local M-estimator and some technical assumptions. The large sample properties for local M-estimator, especially the one-step local M-estimators are collected in Section 3. Section 4 presents the finite sample performance through Monte Carlo simulation study. The estimators are illustrated empirically in Section 5. Some technical lemmas and the proofs for the main theorems are given in Appendix.

2 Local M-estimator and Assumptions

Different from model (1), nonparametric estimators constructed for the coefficients in integrated jump-diffusion model (5) give rise to new challenges for two main reasons.

On the one hand, we usually get observations $\{Y_{i\Delta_n}; i = 1, 2, \cdots\}$ rather than $\{X_{i\Delta_n}; i = 1, 2, \cdots\}$ for model (5). The value of X_{t_i} cannot be obtained from $Y_{t_i} = Y_0 + \int_0^{t_i} X_s ds$ in a fixed sample intervals. Additionally, nonparametric estimations of the unknown qualities in model (5) cannot in principle be constructed on the observations $\{Y_{i\Delta_n}; i = 1, 2, \cdots\}$ due to the unknown conditional distribution of Y. As Nicolau [20] showed, with observations $\{Y_{i\Delta_n}; i = 1, 2, \cdots\}$ and given that

$$Y_{i\Delta_n} - Y_{(i-1)\Delta_n} = \int_{(i-1)\Delta_n}^{i\Delta_n} X_u du,$$

we can obtain an approximation value of $X_{i\Delta_n}$ by

$$\widetilde{X}_{i\Delta_n} = \frac{Y_{i\Delta_n} - Y_{(i-1)\Delta_n}}{\Delta_n}. \tag{6}$$

On the other hand, the Markov properties for statistical inference of unknown qualities in model (5) based on the samples $\{\widetilde{X}_{i\Delta_n}; i = 1, 2, \cdots\}$ should be built, which are infinitesimal conditional expectations characterized by infinitesimal operators as equations (2) and (3). Fortunately, under Lemma 1 we can build the following infinitesimal conditional expectations for model (5)

$$E\left[\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} | \mathscr{F}_{(i-1)\Delta_n}\right] = \mu(X_{(i-1)\Delta_n}) + O_p(\Delta_n), \tag{7}$$

$$E\left[\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})^2}{\Delta_n} | \mathscr{F}_{(i-1)\Delta_n}\right] = \frac{2}{3}\sigma^2(X_{(i-1)\Delta_n}) + \frac{2}{3}\int_{\mathscr{E}} c^2(X_{(i-1)\Delta_n}, z) f(z) dz + O_p(\Delta_n). \tag{8}$$

where $\mathscr{F}_t = \sigma\{X_s, s \le t\}$. One can refer to Appendix A in Song, Lin and Wang [25] for detailed calculations.

Our aim is to estimate the functions $\mu(x)$ and $M(x) := \sigma^2(x) + \int_{\mathscr{E}} c(x, z) f(z) dz$ by $\{\widetilde{X}_{i\Delta_n}; i = 1, 2, \cdots\}$. For the given $\{\widetilde{X}_{i\Delta_n}; i = 1, 2, \cdots\}$, the local linear estimators for $\mu(x)$ and $M(x)$ based on infinitesimal conditional expectations (7) and (8) are defined as the solutions to the following weighted least squares problems: find a_1, b_1, a_2, b_2 to minimize

$$\sum_{i=1}^{n} \Big(\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - a_1 - b_1\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)\Big)^2 K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big), \tag{9}$$

$$\sum_{i=1}^{n} \Big(\frac{\frac{3}{2}(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})^2}{\Delta_n} - a_2 - b_2\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)\Big)^2 K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big), \tag{10}$$

where $K(\cdot)$ is the kernel function and h_n is a sequence of positive numbers, satisfies $h_n \to 0$ as $n \to \infty$.

Obviously, criterions (9) and (10) are based on the least-squares principle and are not robust. To overcome this shortcoming of lack of robustness, the least-squares principle is placed by

$$arg \min_{a_1,b_1} \sum_{i=1}^{n} \rho_1\Big(\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - a_1 - b_1\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big), \tag{11}$$

$$arg \min_{a_2,b_2} \sum_{i=1}^{n} \rho_2\Big(\frac{\frac{3}{2}(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})^2}{\Delta_n} - a_2 - b_2\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big), \tag{12}$$

these are equal to satisfying the following local estimation equations:

$$\Psi_n := \sum_{i=1}^{n} \psi_1\Big(\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - a_1 - b_1\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \begin{pmatrix} 1 \\ \widetilde{X}_{(i-1)\Delta_n} - x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{13}$$

$$\Phi_n := \sum_{i=1}^{n} \psi_2\Big(\frac{\frac{3}{2}(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})^2}{\Delta_n} - a_2 - b_2\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \begin{pmatrix} 1 \\ \widetilde{X}_{(i-1)\Delta_n} - x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{14}$$

where $\rho_1(\cdot)$ and $\rho_2(\cdot)$ are given outlier-resistant functions and $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are the derivatives of $\rho_1(\cdot)$ and $\rho_2(\cdot)$, respectively. The local M-type estimators of $\mu(x)$ is defined as $\hat{\mu}(x) = \hat{a}_1$, which is the solution to equation (13). the local M-type estimators of $M(x)$ is defined as $\hat{M}(x) = \hat{a}_2$, which is the solution to equation (14).

The assumptions of this paper are listed below, which confirm the large sample properties of the constructed estimators based on (13) and (14).

Assumption 1 *i) (Local Lipschitz continuity) For each $n \in \mathbb{N}$, there exist a constant L_n and a function $\zeta_n : \mathscr{E} \to \mathbb{R}_+$ with $\int_{\mathscr{E}} \zeta_n^2(z) f(z) dz < \infty$ such that, for any $|x| \le n, |y| \le n$,*

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \le L_n|x - y|, \quad |c(x,z) - c(y,z)| \le \zeta_n(z)|x - y|.$$

(ii (Linear growthness) For each $n \in \mathbb{N}$, there exist ζ_n as above and C, such that for all $x \in \mathbb{R}$,

$$|\mu(x)| + |\sigma(x)| \le C(1 + |x|), \ |c(x,z)| \le \zeta_n(z)(1 + |x|).$$

Remark 1 *This assumption guarantees the existence and uniqueness of a solution to stochastic differential equation (5), see Jacod and Shiryayev [15].*

Assumption 2 *The process $X = \{X_t\}_{\ge 0}$ is ergodic and stationary with a finite invariant measure $\phi(x)$. For a given point x_0, the stationary probability measure $p(x)$ of the process X is continuous at x_0 and $p(x_0) > 0$. Furthermore, the process X is $\rho-$mixing with $\sum_{i \ge 1} \rho(i\Delta_n) = O(\frac{1}{\Delta_n^{\alpha}})$, $n \to \infty$, where $\alpha < 1$.*

Remark 2 *The hypothesis that X_t is a stationary process is obviously a plausible assumption because for major integrated time series data, a simple differentiation generally assures stationarity. The same condition yielding information on the rate of decay of $\rho-$mixing coefficients for X_t was mentioned the Assumption 3 in Gugushvili and Spereij [11].*

Assumption 3 *The kernel $K(\cdot) : \mathbb{R} \to \mathbb{R}^+$ is a positive, symmetric and continuously differentiable function satisfying:*

$$\int K(u)du = 1, \int uK(u)du = 0,\ K_2 := \int K^2(u)du < \infty.$$

Assumption 4 *For $2 \le i \le n$,*

$$\lim_{h\to 0} E\Big[\ \frac{1}{h}\ |K^{'}(\xi_{n,i})|^{\alpha}\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)^{m}\Big] < \infty$$

where $\alpha = 1, 2$ or 4, $m = 0, 1$ or 2 and $\xi_{n,i} = \theta(\frac{x-X_{(i-1)\Delta_n}}{h})+(1-\theta)(\frac{x-\widetilde{X}_{(i-1)\Delta_n}}{h})$, $0 \le \theta \le 1$.

Remark 3 *As Nicolau [20] pointed out this assumption is generally satisfied under very weak conditions. For instance, with a Gaussian kernel and a Cauchy stationary density (which has heavy tails) we still have $\lim_{h\to 0} E\big[(\frac{1}{h})|K^{'}(\frac{X}{h})|^4\big] < \infty$. Notice that the expectation with respect to the distribution $\xi_{n,i}$ depends on the stationary densities of X and $\widetilde{X}$ because $\xi_{n,i}$ is a convex linear combination of X and $\widetilde{X}$.*

Assumption 5 *For every $p \ge 1$, $\sup_{t\ge 0} E[|X_t|^p] < \infty$, and $\int_{\mathscr{E}} |z|^p f(z)dz < \infty$.*

Remark 4 *If X is a Lévy process with bounded jumps (i.e., $\sup_t |\Delta X_t| \le C < \infty$ almost surely, where C is a nonrandom constant), then $E\{|X_t^n|\} < \infty\ \forall n$, that is, X_t has bounded moments of all orders, see Protter [19]. This condition is widely used in the estimation of an ergodic diffusion or jump-diffusion from discrete observations, see Florens-Zmirou [9], Kessler [18], Shimizu and Yoshida [23].*

Assumption 6 $\Delta_n \to 0$, $(\frac{n\Delta_n}{h_n})(\Delta_n \log(\frac{1}{\Delta_n}))^{\frac{1}{2}} \to 0$, $\frac{\Delta_n \log(1/\Delta_n)}{h_n^2 \Delta_n^{\alpha}} \to 0$, $h_n^5 n\Delta_n = O(1)$ *as* $n \to \infty$.

Assumption 7 *Let $\psi_i(\cdot)$ be the derivative of $\rho_i(\cdot)$ with $i = 1, 2$, $M(x) = \sigma^2(x)+ \int_{\mathscr{E}} c^2(x, z)f(z)dz$, we assume that*

(i) $E\Big[\psi_1(u_{i\Delta_n})|X_{(i-1)\Delta_n} = x\Big] = O(\delta)$ *with* $u_{i\Delta_n} = \frac{\widetilde{X}_{(i+1)\Delta_n}-\widetilde{X}_{i\Delta_n}}{\Delta_n} - \mu(X_{(i-1)\Delta_n})$;

(ii) $E\Big[\psi_1(u^{'}_{i\Delta_n})|X_{(i-1)\Delta_n} = x\Big] = O(\delta)$ *with* $u^{'}_{i\Delta_n} = \frac{X_{i\Delta_n}-X_{(i-1)\Delta_n}}{\Delta_n} - \mu(X_{(i-1)\Delta_n})$;

(iii) $E\Big[\psi_2(v_{i\Delta_n})|X_{(i-1)\Delta_n} = x\Big] = O(\delta)$ *with* $v_{i\Delta_n} = \frac{\frac{3}{2}(\widetilde{X}_{(i+1)\Delta_n}-\widetilde{X}_{i\Delta_n})^2}{\Delta_n} - M(X_{(i-1)\Delta_n})$;

(iv) $E\Big[\psi_2(v^{'}_{i\Delta_n})|X_{(i-1)\Delta_n} = x\Big] = O(\delta)$ *with* $v^{'}_{i\Delta_n} = \frac{(X_{i\Delta_n}-X_{(i-1)\Delta_n})^2}{\Delta_n} - M(X_{(i-1)\Delta_n})$.

Assumption 8 *The function $\psi_i(\cdot)$ is continuous and has a derivative $\psi_i^{'}(\cdot)$ almost everywhere. Furthermore, it is also assumed that*

$$E\Big[\psi_1^{'}(u_{i\Delta_n})|X_{(i-1)\Delta_n}=x\Big]>0,\quad E\Big[\psi_2^{'}(v_{i\Delta_n})|X_{(i-1)\Delta_n}=x\Big]>0,$$

$$E\Big[\psi_1^{2}(u_{i\Delta_n})|X_{(i-1)\Delta_n}=x\Big]>0,\quad E\Big[\psi_2^{2}(v_{i\Delta_n})|X_{(i-1)\Delta_n}=x\Big]>0,$$

$$E\Big[\psi_1^{'2}(u_{i\Delta_n})|X_{(i-1)\Delta_n}=x\Big]>0,\quad E\Big[\psi_2^{'2}(v_{i\Delta_n})|X_{(i-1)\Delta_n}=x\Big]>0,$$

and these functions are continuous at x_0.

Assumption 9 *The function $\psi^{'}(\cdot)$ satisfies*

$$E\Big[\sup_{|z|\le\delta}|\psi_1^{'}(u_{i\Delta_n}+z)-\psi^{'}(u_{i\Delta_n})||X_{(i-1)\Delta_n}=x\Big]=o(1),$$
$$E\Big[\sup_{|z|\le\delta}|\psi_1(u_{i\Delta_n}+z)-\psi(u_{i\Delta_n})-\psi^{'}(u_{i\Delta_n})z||X_{(i-1)\Delta_n}=x\Big]=o(\delta),$$
$$E\Big[\sup_{|z|\le\delta}|\psi_2^{'}(v_{i\Delta_n}+z)-\psi^{'}(v_{i\Delta_n})||X_{(i-1)\Delta_n}=x\Big]=o(1),$$
$$E\Big[\sup_{|z|\le\delta}|\psi_2(v_{i\Delta_n}+z)-\psi(v_{i\Delta_n})-\psi^{'}(v_{i\Delta_n})z||X_{(i-1)\Delta_n}=x\Big]=o(\delta)$$

as $\delta\to 0$ uniformly in x in a neighborhood of x_0.

Remark 5 *The assumptions (7) - (9) imposed on $\psi(\cdot)$ are mild and fulfilled for Huber's $\psi(\cdot)$. For more details about these conditions one can refer to Fan and Jiang [10] such as (A.5), (A.7) and (A.8) in it.*

3 Large sample properties.

We lay out some notations for convenience. Write:

$$K_l=\int K(u)u^l du\ ,\qquad J_l=\int u^l K^2(u)du\ ,\qquad for\quad l\ge 0,$$

$$U\ =\ \begin{pmatrix}K_0 & K_1\\ K_1 & K_2\end{pmatrix}\ ,\qquad V\ =\ \begin{pmatrix}J_0 & J_1\\ J_1 & J_2\end{pmatrix}\ ,\qquad A\ =\ \begin{pmatrix}K_2\\ K_3\end{pmatrix},$$

$$G_1(x)=E\Big[\psi^{'}(u_{i\Delta_n})|X_{(i-1)\Delta_n}=x\Big]\ ,\ G_1^{'}(x)=E\Big[|\psi^{'}(u_{i\Delta_n})||X_{(i-1)\Delta_n}=x\Big]\ ,$$

$$H_1(x)=E\Big[\psi^{'}(v_{i\Delta_n})|X_{(i-1)\Delta_n}=x\Big]\ ,\ H_1^{'}(x)=E\Big[|\psi^{'}(v_{i\Delta_n})||X_{(i-1)\Delta_n}=x\Big]\ ,$$

$$G_2(x)=E\Big[\Delta_n\psi^2(u_{i\Delta_n}^{'})|X_{(i-1)\Delta_n}=x\Big]\ ,\ H_2(x)=E\Big[\Delta_n\psi^2(v_{i\Delta_n}^{'})|X_{(i-1)\Delta_n}=x\Big]\ ,$$

We have the following asymptotic results.

Theorem 1 *Under the Assumptions 1-9, then there exist solutions, denoted by $\hat{\mu}_n(x_0)$ and $\hat{\mu}_n^{'}(x_0)$ to equation (13), and there exist solutions, denoted by $\hat{M}_n(x_0)$ and $\hat{M}_n^{'}(x_0)$ to equation (14) such that*

(i) $\begin{pmatrix} \hat{\mu}_n(x_0) - \mu(x_0) \\ h(\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)) \end{pmatrix} \xrightarrow{P} 0, \qquad$ *as* $n \longrightarrow \infty$

(ii)

$$\sqrt{nh\Delta_n}\Big[\begin{pmatrix} \hat{\mu}_n(x_0) - \mu(x_0) \\ h(\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)) \end{pmatrix} - \frac{h^2\mu^{''}(x_0)}{2}U^{-1}A\Big] \xrightarrow{D} N(0, \Sigma_2),$$

where $\Sigma_2 = \frac{G_2(x_0)}{G_1^2(x_0)p(x_0)}U^{-1}VU^{-1}$,

(iii) $\begin{pmatrix} \hat{M}_n(x_0) - M(x_0) \\ h(\hat{M}_n^{'}(x_0) - M^{'}(x_0)) \end{pmatrix} \xrightarrow{P} 0, \qquad$ *as* $n \longrightarrow \infty$

(iv)

$$\sqrt{nh\Delta_n}\Big[\begin{pmatrix} \hat{M}_n(x_0) - M(x_0) \\ h(\hat{M}_n^{'}(x_0) - M^{'}(x_0)) \end{pmatrix} - \frac{h^2 M^{''}(x_0)}{2}U^{-1}A\Big] \xrightarrow{D} N(0, \Sigma_4),$$

where $\Sigma_4 = \frac{H_2(x_0)}{H_1^2(x_0)p(x_0)}U^{-1}VU^{-1}$.

A non-iterative one-step local M-estimator with a similar performance as local linear M-estimator is constructed through Newton's procedure, which is following the idea of Bickel [5], Fan and Jiang [9]. Following is an outline of the procedure.

Solve the nonlinear equations (13), (14) by Newton's method for the first iteration with the initial value $a_{10} = \hat{\mu}_0(x)$ and $b_{10} = \hat{\mu}_0^{'}(x)$, $a_{20} = \hat{M}_0(x)$ and $b_{20} = \hat{M}_0^{'}(x)$, which can be derived in the simple and explicit expression. Then the form for the first iteration of Newton's method is as follows

$$\begin{pmatrix} \widetilde{\mu}_n(x) \\ \widetilde{\mu}_n^{'}(x) \end{pmatrix} = \begin{pmatrix} \hat{\mu}_0(x) \\ \hat{\mu}_0^{'}(x) \end{pmatrix} - W_n^{-1}\Psi_n(\hat{\mu}_0(x), \hat{\mu}_0^{'}(x)), \tag{15}$$

where

$$W_n = \begin{pmatrix} \frac{\partial}{\partial a_1}\Psi_{n1}(\hat{\mu}_0(x), \hat{\mu}_0^{'}(x)), & \frac{\partial}{\partial b_1}\Psi_{n1}(\hat{\mu}_0(x), \hat{\mu}_0^{'}(x)) \\ \frac{\partial}{\partial a_1}\Psi_{n2}(\hat{\mu}_0(x), \hat{\mu}_0^{'}(x)), & \frac{\partial}{\partial b_1}\Psi_{n2}(\hat{\mu}_0(x), \hat{\mu}_0^{'}(x)) \end{pmatrix},$$

and

$$\begin{pmatrix} \widetilde{M}_n(x) \\ \widetilde{M}_n^{'}(x) \end{pmatrix} = \begin{pmatrix} \hat{M}_0(x) \\ \hat{M}_0^{'}(x) \end{pmatrix} - Z_n^{-1}\Psi_n(\hat{M}_0(x), \hat{M}_0^{'}(x)), \tag{16}$$

where

$$Z_n = \begin{pmatrix} \frac{\partial}{\partial a_2}\Phi_{n1}(\hat{M}_0(x), \hat{M}_0^{'}(x)), & \frac{\partial}{\partial b_2}\Phi_{n1}(\hat{M}_0(x), \hat{M}_0^{'}(x)) \\ \frac{\partial}{\partial a_2}\Phi_{n2}(\hat{M}_0(x), \hat{M}_0^{'}(x)), & \frac{\partial}{\partial b_2}\Phi_{n2}(\hat{M}_0(x), \hat{M}_0^{'}(x)) \end{pmatrix}.$$

The estimators $\widetilde{\mu}_n(x)$, $\widetilde{\mu}_n^{'}(x)$, $\widetilde{M}_n(x)$ and $\widetilde{M}_n^{'}(x)$ are called "one-step local M-estimators". Subsequently, we will show that the one-step local M-estimators

possess the same asymptotic performance as the M-estimator as long as the initial estimators $\widetilde{\mu}_n(x)$, $\widetilde{\mu}_n^{'}(x)$, $\widetilde{M}_n(x)$ and $\widetilde{M}_n^{'}(x)$ satisfy mild conditions, which to some extend reduce computational cost without loss of their property.

Theorem 2 *Assume that the initial estimators satisfy*

$$\hat{\mu}_0(x_0)-\mu(x_0) = O_p\Big(h_n^2+\frac{1}{\sqrt{nh_n\Delta_n}}\Big) \quad and \quad h_n\Big(\hat{\mu}_0^{'}(x_0)-\mu^{'}(x_0)\Big) = O_p\Big(h_n^2+\frac{1}{\sqrt{nh_n\Delta_n}}\Big),$$

$$\hat{M}_0(x_0)-M(x_0) = O_p\Big(h_n^2+\frac{1}{\sqrt{nh_n\Delta_n}}\Big) \quad and \quad h_n\Big(\hat{M}_0^{'}(x_0)-M^{'}(x_0)\Big) = O_p\Big(h_n^2+\frac{1}{\sqrt{nh_n\Delta_n}}\Big).$$

Then, under the Assumptions 1-9, then we have

$$\sqrt{nh\Delta_n}\Big[\begin{pmatrix}\widetilde{\mu}_n(x_0)-\mu(x_0)\\ h(\widetilde{\mu}_n^{'}(x_0)-\mu^{'}(x_0))\end{pmatrix} - \frac{h^2\mu^{''}(x_0)}{2}U^{-1}A\Big] \xrightarrow{D} N(0,\Sigma_2),$$

where $\Sigma_2 = \frac{G_2(x_0)}{G_1^2(x_0)p(x_0)}U^{-1}VU^{-1}$, *and*

$$\sqrt{nh\Delta_n}\Big[\begin{pmatrix}\widetilde{M}_n(x_0)-M(x_0)\\ h(\widetilde{M}_n^{'}(x_0)-M^{'}(x_0))\end{pmatrix} - \frac{h^2M^{''}(x_0)}{2}U^{-1}A\Big] \xrightarrow{D} N(0,\Sigma_4),$$

where $\Sigma_4 = \frac{H_2(x_0)}{H_1^2(x_0)p(x_0)}U^{-1}VU^{-1}$.

Remark 6 *In contrary to the second-order diffusion model without jumps (Nicolau [20]), the second infinitesimal moment estimator has a rate of convergence that is the same as the rate of convergence of the first infinitesimal moment estimator. Apparently, this is due to the presence of discontinuous breaks that have an equal impact on all the functional estimates. As Johannes [17] pointed out, for the conditional variance of interest rate changes, not only diffusion play a certain role, but also jumps account for more than half at lower interest level rates, almost two-thirds at higher interest level rates, which dominate the conditional volatility of interest rate changes. Thus, it is extremely important to estimate the conditional variance as* $\sigma^2(x) + \int_{\mathscr{E}} c^2(x,z)f(z)dz$ *which reflects the fluctuation of the underlying asset.*

Remark 7 *It is very important to consider the choice of the bandwidth in nonparametric estimation. Here we will select the optimal bandwidth* h_n *based on the mean squared error (MSE) and the asymptotic theory in Theorem 1 and 2. Take* $\mu(x)$ *for example, the optimal smoothing parameter* h_n *for local M-estimator of* $\mu(x)$ *is given that*

$$h_{n,opt,fi} = \left(\frac{1}{n\Delta_n}\cdot\frac{4}{\mu^{2''}(x)a_1}\right)^{\frac{1}{5}} = O_p\left(\frac{1}{n\Delta_n}\right)^{\frac{1}{5}},$$

where a_1 *denotes the first elements in the vector* $U^{-1}A$. *Meanwhile, the optimal smoothing parameter* h_n *for one-step local M-estimator of* $\mu(x)$ *is the same as that for local M-estimator as far as the initial estimator is well behaves, such as the local linear estimator. The optimal bandwidth coincides with that in Bandi and Nguyen [2].*

4 Monte Carlo Simulation Study

In this section, we conduct a simple Monte Carlo simulation experiment aimed at the small-sampling performance of the local M-estimators constructed as (13) and the one-step local M-estimators constructed as (15) . Assessment will be made between them and local linear estimators by comparing their mean square error (MSE). Our experiment is based on the following model:

$$\begin{cases} dY_t = X_{t-}dt, \\ dX_t = -10X_{t-}dt + \sqrt{0.1 + 0.1X_{t-}^2}dW_t + dJ_t, \end{cases} \tag{17}$$

where the coefficients of continuous part equal to the ones used in Nicolau ([20]) and J_t is a compound Poisson jump process, that is, $J_t = \sum_{n=1}^{N_t} Z_{t_n}$ with arrival intensity $\lambda = 20$ and jump size $Z_n \sim \mathscr{N}(0, 0.036^2)$ corresponding to Bandi and Nguyen ([2]), where t_n is the nth jump of the Poisson process N_t. By taking the integral from 0 to t in the second expression of (17), we obtain

$$X_t = -10\int_0^t X_{s-}ds + \int_0^t \sqrt{0.1 + 0.1X_{s-}^2}dW_t + \sum_{n=1}^{N_t} Z_{t_n}. \tag{18}$$

Then we have

$$Y_t = \int_0^t X_{s-}ds = -\frac{1}{10}\Big(X_t - \int_0^t \sqrt{0.1 + 0.1X_{s-}^2}dW_t - \sum_{n=1}^{N_t} Z_{t_n}\Big). \tag{19}$$

X_t can be sampled by the Euler-Maruyama scheme according to (18), which will be detailed in the following algorithm (one can refer to Cont and Tankov [6]).

Algorithm 1 Simulation for trajectories of second-order jump-diffusion model

---

Procedures:

Step 1: generate a standard normal random variate V and transform it into $D_i = \sqrt{0.1 + 0.1X_{t_{i-1}}^2} * \sqrt{\Delta t_i} * V$, where $\Delta t_i = t_i - t_{i-1} = \frac{T}{n}$ is the observation time frequency;

Step 2: generate a Poisson random variate N with intensity $\lambda = 20$;

Step 3: generate N random variables τ_i uniformly distributed in $[0, T]$, which correspond the jump times;

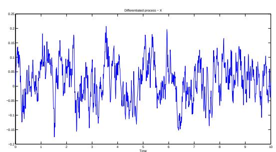
Step 4: generate N random variables $Z_{\tau_i} \sim \mathscr{N}(0, 0.036^2)$, which correspond the jump sizes;

One trajectory for X_t is

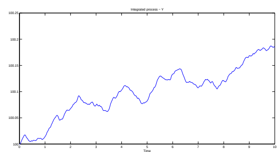
$$X_{t_i} = X_{t_{i-1}} - 10X_{t_{i-1}} * \Delta t_i + D_i + 1_{\{t_{i-1} \le \tau_i < t_i\}} * Z_{\tau_i}.$$

Step 5: By substitution of X_{t_i} in (19), Y_{t_i} can be sampled.

One sample trajectory of the differentiated process X_t and integrated process Y_t with $T = 10,\ n = 1000,\ X_0 = 0$ and $Y_0 = 100$ using Algorithm 1 is shown in FIG 1. Through observation on FIG 1(b), we can find the following features of the integrated process Y_t: absent mean-reversion, persistent shocks, time-dependent mean and variance, nonnormality, etc.



(a) The differentiated process X_t



(b) The integrated process Y_t

Figure 1: the Sample Paths of X_t and Y_t

Throughout this section, we take $\rho(u) = \begin{cases} u^2/2, & |u| \leq 0.135 \\ 0.135|u| - 0.135^2/2, & |u| > 0.135 \end{cases}$, which deduces the Huber's function $\psi(u) = \rho^{'}(u) = \max\{-0.135, \min\{u, 0.135\}\}$. We will consider various lengths of observation time interval T $(= 50, 100, 500)$ and sample sizes n $(= 500, 1000, 5000)$ with $\Delta_n = \frac{T}{n}$. We use Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and the common bandwidth $h = c\hat{S}(n\Delta_n)^{-\frac{1}{5}} = c\hat{S}T^{-\frac{1}{5}}$, where $\hat{S}$ denotes the standard deviation of the data and c represents different constants for different estimators with c = 2.8 for $\hat{\mu}_n(x)$ and c = 1.3 for $\hat{M}_n(x)$. Since Bandi and Nguyen [2] and Theorems 1 - 2 above show that the convergence rates of estimators depend on the time span $n\Delta_n = T$, the character of the bandwidth h is used in terms of T instead of the sampling size n, which is different from the continuous case in Nicolau ([20]) (One can refer to part (4.2) in Xu and Phillips [30] for more details).

Take the estimator $\hat{\mu}_n(x)$ of $\mu(x)$ for example to demonstrate the small-sampling performance. Due to no explicit expression for $\hat{\mu}_n(x)$, we should deduce $\hat{\mu}_n(x)$ by iteration based on any initial value $\hat{\mu}_0(x)$ and the following iterative relation:

$$\begin{pmatrix} \hat{\mu}_n(x) \\ \hat{\mu}_n^{'}(x) \end{pmatrix} = \begin{pmatrix} \hat{\mu}_{n-1}(x) \\ \hat{\mu}_{n-1}^{'}(x) \end{pmatrix} - W_n^{-1}(\hat{\mu}_{n-1}(x), \hat{\mu}_{n-1}^{'}(x))\Psi_n(\hat{\mu}_{n-1}(x), \hat{\mu}_{n-1}^{'}(x)), \quad (20)$$

where $\hat{\mu}_{n-1}(x)$ and $\hat{\mu}_{n-1}^{'}(x)$ are the $n-$th iteration and as for W_n and Ψ_n, one can refer to equations (13) and (15). The loop termination criterion are $n = 600$ or

$$\left|\left| \begin{pmatrix} \hat{\mu}_n(x) \\ \hat{\mu}_n^{'}(x) \end{pmatrix} - \begin{pmatrix} \hat{\mu}_{n-1}(x) \\ \hat{\mu}_{n-1}^{'}(x) \end{pmatrix} \right|\right| \leq 10^{-4}. \quad (21)$$

Figure 2 represents the local linear estimator, the one-step local M-estimator and the fully iterative local-M estimator for $\mu(x)$ from a sample with $T = 10$ and $n = 1000$, which shows the one-step or fully iterative local-M estimators performs a little better than local linear estimator, especially on the boundary.

Next, we will assess the performance of the fully iterative local M-estimator, the one-step local M-estimator in this paper and local linear estimator via the Mean Square Errors (MSE)

$$MSE = \frac{1}{m} \sum_{k=0}^{m} \left\{ \hat{\mu}(x_k) - \mu(x_k) \right\}^2 , \tag{22}$$

where $\hat{\mu}(x)$ is the estimator of $\mu(x)$ and $\{x_k\}_1^m$ are chosen uniformly to cover the range of sample path of X_t. Table 1 gives the results on the MSE of local linear estimator (MSE-LL), the one-step local-M estimator (MSE-OS) and the fully iterative local-M estimator (MSE-FI) for the drift function $\mu(x)$ with jump size $Z_n \sim \mathscr{N}(0, 0.036^2)$ over 100 replicates. Table 2 reports the results on MSE-LL, MSE-OS and MSE-FI for the drift function $\mu(x)$ with different types of jump size Z_n over 100 replicates.

We can notice that the one-step or fully iterative local-M estimator performs a little better than the local linear estimator in terms of the MSE from Table 1 and 2. From Table 1, we can get the other three findings. Firstly, for the same time interval T, as the sample sizes n tends larger, the performances of these estimators improved. Secondly, for the same sample sizes n, as the time interval T expands larger, the performances of these estimators get worse due to the fact that more jumps happens in larger time interval T in steps 3 of Algorithm 1. Thirdly, the one-step local-M estimator performs almost as the fully iterative local-M estimator does, which to some extent reduces computational cost. From Table 2, we can know that the performances of the one-step or fully iterative local-M estimator are more robust than the local linear estimator, especially when the jump size $Z_n \sim \mathscr{N}(0, 1)$ or $Z_n \sim Cauchy(0, 1)$. In addition, the one-step local-M estimator performs almost as the fully iterative local-M estimator does, which confirms the result in Theorem 2.

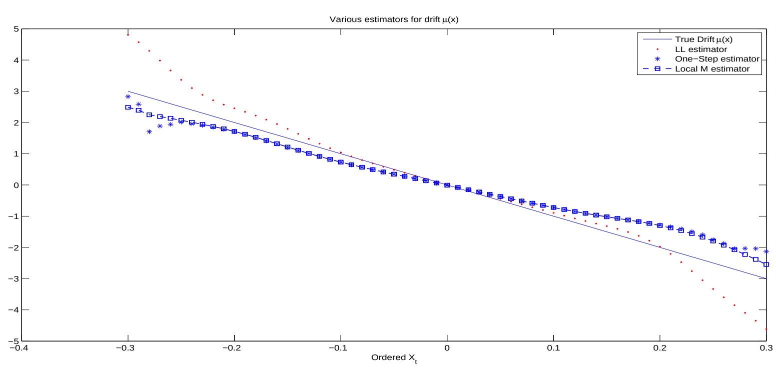


Figure 2: Various Estimators for $\mu(t)$

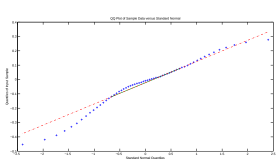
Figure 3 gives the QQ Plots for the one-step and fully iterative local-M estimators of the drift function $\mu(x)$ with $T = 10$ and $\Delta_n = 0.01$. This reveals

Table 1: Simulation results on MSE-LL, MSE-OS and MSE-LM for three lengths of time interval (T) and three sample sizes for $\mu(x) = -10x$ with jump size $Z_n \sim \mathscr{N}(0, 0.036^2)$ over 100 replicates.

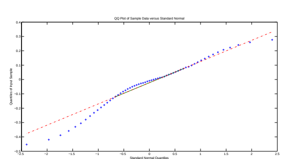
T		$n = 500$	$n = 1000$	$n = 5000$
10	MSE-LL	0.8913	0.2478	0.0591
	MSE-OS	0.8175	0.2261	0.0512
	MSE-FI	0.8040	0.2259	0.0508
50	MSE-LL	2.6146	0.7495	0.0437
	MSE-OS	2.6097	0.7433	0.0422
	MSE-FI	2.6093	0.7430	0.0421
100	MSE-LL	12.4987	3.3573	0.1422
	MSE-OS	12.4303	3.3461	0.1393
	MSE-FI	12.4273	3.3450	0.1393

Table 2: Simulation results on MSE-LL, MSE-OS and MSE-LM for three types of jump size Z_n and three sample sizes for $\mu(x) = -10x$ with $T = 10$ over 100 replicates.

T		$n = 500$	$n = 1000$	$n = 5000$
$Z_n \sim \mathscr{N}(0, 0.036^2)$	MSE-LL	0.8913	0.2478	0.0591
	MSE-OS	0.8175	0.2261	0.0512
	MSE-FI	0.8040	0.2259	0.0508
$Z_n \sim \mathscr{N}(0, 1)$	MSE-LL	0.8570	0.6834	0.1975
	MSE-OS	0.5221	0.3915	0.0639
	MSE-FI	0.4556	0.3887	0.0638
$Z_n \sim Cauchy(0, 1)$	MSE-LL	7.4400	5.8410	2.7078
	MSE-OS	2.9864	0.5740	0.4145
	MSE-FI	2.9813	0.3858	0.3457



(a) QQ plot for One-step local-M estimators



(b) QQ plot for Fully iterative local-M estimators

Figure 3: QQ Plots of Sample Data versus Standard Normal for One-step and Fully iterative local-M estimators of $\mu(x)$

the normality of one-step and fully iterative local-M estimators in finite sample, which confirms the results in Theorems 1 - 2.

5 Empirical Analysis.

In this section, we apply the integrated jump-diffusion to model the stock index of Shanghai Stock Exchange by using the daily (from Jan 4, 2000 to Feb 7, 2017) and one-minute (May 4, 2015 to Aug 31, 2015) data of closing stock price, and then apply the one-step local M-estimators to estimate the unknown the coefficients in model (5). To our knowledge, it is the first article regarding the integrated jump-diffusion model for stock index, under low-frequency data containing the year 2008 (financial crisis), and one-minute high-frequency data.

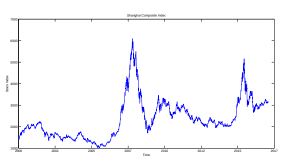
We assume that

$$\begin{cases} d\log Y_t = X_t dt, \\ dX_t = \mu(X_{t-})dt + \sigma(X_{t-})dW_t + \int_{\mathscr{E}} c(X_{t-}, z) r(\omega, dt, dz), \end{cases} \tag{23}$$

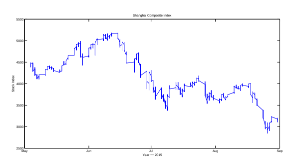
where $\log Y_t$ is the log integrated process for stock index and X_t is the latent process for the log-returns. According to (6), we can get the proxy of the latent process

$$\widetilde{X}_{i\Delta_n} = \frac{\log Y_{i\Delta_n} - \log Y_{(i-1)\Delta_n}}{\Delta_n}. \tag{24}$$

The plots of the stock index (SSE) and its proxy in daily and one-minute frequency data are shown in Figure 4 and 5.

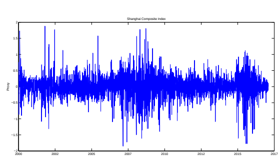


(a) Daily Data from Jan 4, 2000 to Feb 7, 2017

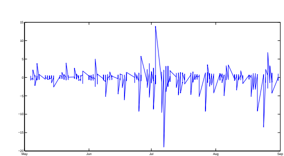


(b) One-Minute Data from May 4, 2015 to Aug 31, 2015

Figure 4: the Time Series of the Stock Index (SSE) in Daily and One-Minute Frequency Data



(a) Daily Proxy from Jan 4, 2000 to Feb 7, 2017



(b) One-Minute Proxy from May 4, 2017 to Aug 31, 2017

Figure 5: the Proxy of the Stock Index (SSE) in Daily and One-Minute Frequency Data

First, we test the existence of jumps for the proxy X_t through the test statistic proposed in Barndorff-Nielsen and Shephard [4]. For daily data, the

value of the test statistic is -6.1721, which exceeds [-1.96, 1.96]. For one-minute data, the value of the test statistic is 6.9171, which also exceeds [-1.96, 1.96]. We can obtain that there exists jumps in daily or one-minute data at the 5% significance level, which coincides with macroeconomic shocks such as 2008's financial crisis throughout the world, 2015's disaster in China stock market etc, and confirms the validity of model (5) not model (4) for the stock index in Shanghai Stock Exchange by the integrated process.

Then, we test the stationarity for Y_t and the proxy X_t through the Augmented Dickey-Fuller test statistic.

Table 3: Augmented Dickey-Fuller stationarity test.

Data		TestStat	CriticalValue	PValue
Daily data	Y_t	-1.7127	-2.8635	0.4237
	X_t	-62.7877	-2.8635	¡0.01
One-Minute data	Y_t	-0.1709	-2.861	0.9395
	X_t	-103.9539	-2.861	¡0.01

From table 3, we see that the null hypothesis of non-stationarity is accepted at the 5% significance level for the stock index Y_t, but is rejected for the proxy of X_t, which confirms the assumption of stationary by differencing.

Finally, we will employ the one-step local M-estimator (15) and (16) to estimate the unknown qualities $\mu(x)$ and $M(x)$ based on (24) with $\Delta_n = \frac{1}{20}$ for daily data ($t = 1$ meaning one month) and $\Delta_n = \frac{1}{242}$ for one-minute data ($t = 1$ meaning one day). Meanwhile, the 10% observations are used for the initial estimator for (20). The estimation curves for unknown qualities in daily and one-minute frequency data are demonstrated in Figure 6 and 7.

(a) One-step local M-estimator of the drift coefficient in daily data

(b) One-step local M-estimator of the drift coefficient in one-minute data

Figure 6: One-step local M-estimator of the drift coefficient for the Stock Index (SSE) in Daily and One-Minute Frequency Data

It can be shown that the linear shape with negative coefficient for drift estimator in FIG 6(a) & 7(a) which indicates that the higher log-return increments correspond to the lower drift in the latent process, (this fact coincides with the economic phenomenon of mean reversion) and the quadratic form with positive

(a) One-step local M-estimator of the volatility coefficient in daily data

(b) One-step local M-estimator of the volatility coefficient in one-minute data

Figure 7: One-step local M-estimator of the volatility coefficient for the Stock Index (SSE) in Daily and One-Minute Frequency Data

coefficient for volatility estimator in FIG 6(b) & 7(b) which reveals that the higher absolute value of log-return increments correspond to the higher volatility in the latent process (this fact coincides with the economic phenomenon of volatility smile).

6 Appendix.

In this section, we first present some technical lemmas and the proofs for the main theorems.

6.1 Some Technical Lemmas with Proofs

Lemma 1 *(Shimizu and Yoshida [23]) Let Z be a d-dimensional solution-process to the stochastic differential equation*

$$Z_t = Z_0 + \int_0^t \mu(Z_{s-})ds + \int_0^t \sigma(Z_{s-})dW_s + \int_0^t \int_{\mathscr{E}} c(Z_{s-}, z) r(\omega, dt, dz),$$

where Z_0 is a random variable, $\mathscr{E} = \mathbb{R}^d \backslash \{0\}$, $\mu(x), c(x,z)$ are d-dimensional vectors defined on $\mathbb{R}^d, \mathbb{R}^d \times \mathscr{E}$ respectively, $\sigma(x)$ is a $d \times d$ diagnonal matrix defined on $\mathbb{R}^d$, and W_t is a d-dimensional vector of independent Brownian motions.

Let g be a $C^{2(l+1)}$-class function whose derivatives up to $2(l+1)$th are of polynomial growth. Assume that the coefficient $\mu(x), \sigma(x)$, and $c(x,z)$ are C^{2l}-class function whose derivatives with respective to x up to $2l$th are of polynomial growth. Under Assumption 6, the following expansion holds

$$E[g(Z_t)|\mathscr{F}_s] = \sum_{j=0}^{l} L^j g(Z_s) \frac{\Delta_n^j}{j!} + R, \tag{25}$$

for $t > s$ and $\Delta_n = t - s$, where $R = \int_0^{\Delta_n} \int_0^{u_1} \cdots \int_0^{u_l} E[L^{l+1} g(Z_{s+u_{l+1}})|\mathscr{F}_s] du_1 \ldots du_{l+1}$

is a stochastic function of order Δ_n^{l+1}, $Lg(x) = \partial_x^* g(x)\mu(x) + \frac{1}{2} tr[\partial_x^2 g(x)\sigma(x)\sigma^*(x)] + \int_{\mathscr{E}} \{g(x + c(x,z)) - g(x) - \partial_x^* g(x)c(x,z)\} f(z)dz$.

Remark 8 *Consider a particularly important model:*

$$\begin{cases} dY_t = X_{t-}dt, \\ dX_t = \mu(X_{t-})dt + \sigma(X_{t-})dW_t + \int_{\mathscr{E}} c(X_{t-}, z)r(w, dt, dz). \end{cases}$$

As $d = 2$, *we have*

$$\begin{aligned} Lg(x,y) &= x(\partial g/\partial y) + \mu(x)(\partial g/\partial x) + \tfrac{1}{2}\sigma^2(x)(\partial^2 g/\partial x^2) \\ &\quad + \int_{\mathscr{E}} \{g(x + c(x,z), y) - g(x,y) - \tfrac{\partial g}{\partial x} \cdot c(x,z)\} f(z)dz. \end{aligned} \tag{26}$$

Lemma 2 *[Jacod [14]] A sequence of* $\mathbb{R}-$*valued variables* $\{\zeta_{n,i} : i \geq 1\}$ *defined on the filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbb{P})$ *is* $\mathcal{F}_{i\Delta_n}-$*measurable for all* n, i. *Assume there exists a continuous adapted* $\mathbb{R}-$*valued process of finite variation* B_t *and a continuous adapted and increasing process* C_t, *for any* $t > 0$, *we have*

$$\sup_{0\leq s\leq t} \Big| \sum_{i=1}^{[s/\Delta_n]} \mathbb{E}\big[\zeta_{n,i} | \mathcal{F}_{(i-1)\Delta_n}\big] - B_s \Big| \xrightarrow{P} 0, \tag{27}$$

$$\sum_{i=1}^{[t/\Delta_n]} \left(\mathbb{E}\big[\zeta_{n,i}^2 | \mathcal{F}_{(i-1)\Delta_n}\big] - \mathbb{E}^2\big[\zeta_{n,i} | \mathcal{F}_{(i-1)\Delta_n}\big]\right) - C_t \xrightarrow{P} 0, \tag{28}$$

$$\sum_{i=1}^{[t/\Delta_n]} \mathbb{E}\big[\zeta_{n,i}^4 | \mathcal{F}_{(i-1)\Delta_n}\big] \xrightarrow{P} 0. \tag{29}$$

Then the processes

$$\sum_{i=1}^{[t/\Delta_n]} \zeta_{n,i} \Rightarrow B_t + M_t,$$

where M_t *is a continuous process defined on the filtered probability space* $(\Omega, P, \mathcal{F})$ *and which, conditionally on the the* $\sigma-$*filter* $\mathcal{F}$, *is a centered Gaussian* $\mathbb{R}-$*valued process with* $E\big[M_t^2 | \mathcal{F}\big] = C_t$.

Remark 9 *Condition (29) is a conditional Lindeberg theorem or Lyapounov's condition, whose aims is to ensure that the limiting process is continuous. It is a particular case of Theorem IX.7.28 in Jacod and Shiryayev [15].*

Lemma 3 *Under the Assumptions 1-9, for any random sequence* $\{\eta_{i\Delta_n}\}_{i=1}^n$ *with* $\max_{1\ \leq i\ \leq n} |\eta_{i\Delta_n}| = o_p(1)$ *and* $l = 0, 1, 2$, *we have*

$$(i)\quad \sum_{i=1}^n \psi_1^{'}(u_{i\Delta_n} + \eta_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)^l = nh^{l+1} G_1(x_0)p(x_0)k_l(1 + o_p(1)),$$

$$(ii)\quad \sum_{i=1}^n \psi_1^{'}(u_{i\Delta_n} + \eta_{i\Delta_n}) R_1\big(\widetilde{X}_{(i-1)\Delta_n}\big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)^l$$

$$= nh^{l+3} \frac{G_1(x_0)}{2} \mu^{''}(x_0)p(x_0)k_{l+2}(1 + o_p(1)),$$

where $R_1\big(\widetilde{X}_{(i-1)\Delta_n}\big) = \mu(\widetilde{X}_{(i-1)\Delta_n}) - \mu(x_0) - \mu^{'}(x_0)(\widetilde{X}_{(i-1)\Delta_n} - x)$.

Proof. For (i), we have

$$
\begin{aligned}
& \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n} + \eta_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)^l \\
= \; & \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)^l \\
& + \sum_{i=1}^{n} \big[\psi_1^{'}(u_{i\Delta_n} + \eta_{i\Delta_n}) - \psi_1^{'}(u_{i\Delta_n})\big] K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)^l \\
:= \; & T_{n1} + T_{n2}.
\end{aligned}
$$

For T_{n1},

$$
\begin{aligned}
\frac{1}{nh^{l+1}} T_{n1} \; := \; & \frac{1}{nh^{l+1}} \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)^l \\
= \; & \frac{1}{nh^{l+1}} \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n}) K\Big(\frac{X_{(i-1)\Delta_n} - x}{h}\Big)\big(X_{(i-1)\Delta_n} - x\big)^l \\
& + \frac{1}{nh^{l+1}} \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n}) \Big[K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)^l \\
& - K\Big(\frac{X_{(i-1)\Delta_n} - x}{h}\Big)\big(X_{(i-1)\Delta_n} - x\big)^l\Big] \\
:= \; & \frac{1}{nh^{l+1}} T_{n1}^{'} + \delta_n.
\end{aligned}
$$

For $\frac{1}{nh^{l+1}} T_{n1}^{'}$, we only need to show

$$
\begin{aligned}
E\big[\frac{1}{nh^{l+1}} T_{n1}^{'}\big] \; = \; & \frac{1}{nh^{l+1}} \sum_{i=1}^{n} E\big[\psi_1^{'}(u_{i\Delta_n}) K\Big(\frac{X_{(i-1)\Delta_n} - x}{h}\Big)\big(X_{(i-1)\Delta_n} - x\big)^l\big] \\
= \; & \frac{1}{nh^{l+1}} \sum_{i=1}^{n} E\Big\{E\big[\psi_1^{'}(u_{i\Delta_n}) K\Big(\frac{X_{(i-1)\Delta_n} - x}{h}\Big)\big(X_{(i-1)\Delta_n} - x\big)^l | X_{(i-1)\Delta_n} = x_0\big]\Big\} \\
= \; & \frac{1}{nh^{l+1}} \sum_{i=1}^{n} E\Big\{K\Big(\frac{X_{(i-1)\Delta_n} - x}{h}\Big)\big(X_{(i-1)\Delta_n} - x\big)^l E\big[\psi_1^{'}(u_{i\Delta_n}) | X_{(i-1)\Delta_n} = x_0\big]\Big\} \\
= \; & G_1(x_0) \frac{1}{nh^{l+1}} \sum_{i=1}^{n} E\big[K\Big(\frac{X_{(i-1)\Delta_n} - x}{h}\Big)\big(X_{(i-1)\Delta_n} - x\big)^l\big] \\
= \; & G_1(x_0) p(x_0) K_l,
\end{aligned}
$$

by using the assumption on $G_1(x_0)$ and the stationary density of the process X_t, so we have

$$
\frac{1}{nh^{l+1}} \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n}) K\Big(\frac{X_{(i-1)\Delta_n} - x}{h}\Big)\big(X_{(i-1)\Delta_n} - x\big)^l = G_1(x_0) p(x_0) K_l (1 + o_p(1)).
$$

Now we mention a property of uniform boundedness of the increments between $\widetilde{X}_{(i-1)\Delta_n}$ and $X_{(i-1)\Delta_n}$:

$$\begin{aligned}
& \max_{1\leq i\leq n} \big|\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n}\big| \\
\leq \;& \max_{1\leq i\leq n} \frac{1}{\Delta_n}\Big| \int_{(i-2)\Delta_n}^{(i-1)\Delta_n} (X_{s-} - X_{(i-1)\Delta_n})ds\Big| \\
\leq \;& \max_{1\leq i\leq n} \sup_{(i-2)\Delta_n\leq s\leq (i-1)\Delta_n} \big|X_{s-} - X_{(i-1)\Delta_n}\big| \\
= \;& O_{a.s.}(\sqrt{\Delta_n \log(1/\Delta_n)}), \qquad (30)
\end{aligned}$$

For δ_n,

$$\begin{aligned}
\delta_n \;=\; & \frac{1}{nh^{l+1}}\sum_{i=1}^n \psi_1^{'}(u_{i\Delta_n})\Big[K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l - K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\big(X_{(i-1)\Delta_n}-x\big)^l\Big] \\
= \;& \frac{1}{nh^{l+1}}\sum_{i=1}^n \psi_1^{'}(u_{i\Delta_n})\Big[K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l - K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l\Big] \\
+ \;& \frac{1}{nh^{l+1}}\sum_{i=1}^n \psi_1^{'}(u_{i\Delta_n})\Big[K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l - K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\big(X_{(i-1)\Delta_n}-x\big)^l\Big] \\
:= \;& \delta_{1n}+\delta_{2n}.
\end{aligned}$$

Here we only prove $\delta_{1n} \to 0$ for simplicity, the same procedure can also be applied to $\delta_{2n} \to 0$.

$$\begin{aligned}
E|\delta_{1n}| \;\leq\; & \frac{1}{nh^{l+1}}\sum_{i=1}^n E\Big|\psi_1^{'}(u_{i\Delta_n})\Big[K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big) - K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big]\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l\Big| \\
= \;& \frac{1}{nh^{l+1}}\sum_{i=1}^n E\Big\{\big|\Big[K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big) - K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big]\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l \\
& \times\big|E\big[|\psi_1^{'}(u_{i\Delta_n})|\big|X_{(i-1)\Delta_n}=x_0\big]\Big\} \\
= \;& G_1^{'}(x_0)\frac{1}{h^{l+1}}E\Big\{\big|K^{'}(\xi_{ni})\frac{\widetilde{X}_{(i-1)\Delta_n}-X_{(i-1)\Delta_n}}{h_n}\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l\big|\Big\} \\
\leq \;& G_1^{'}(x_0)\cdot\frac{\sqrt{\Delta_n\log(1/\Delta_n)}}{h_n}E\big[\frac{1}{h_n}\big|K^{'}(\xi_{ni})\big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h_n}\big)^l\big|\big] \longrightarrow 0
\end{aligned}$$

by mean-value theorem for $K(\cdot)$, (30), the assumption on $G_1(x_0)$, the stationary density of the process X_t and assumption 5 with $m=l$.

For T_{n2}, for any given $\eta>0$, let $\Lambda_n=(\delta_1,\delta_1,\cdots,\delta_n)^T$,

$$D_\eta = \{\Lambda_n : |\delta_j|\leq \eta, \forall j\leq n\},$$

$$Y(\Lambda_n)=\frac{1}{nh^{l+1}}\sum_{i=1}^n\big[\psi_1^{'}(u_{i\Delta_n}+\delta_j)-\psi_1^{'}(u_{i\Delta_n})\big]K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l.$$

Then

$$\sup_{D_\eta}|Y(\Lambda_n)| \leq \frac{1}{nh^{l+1}}\sum_{i=1}^n \sup_{D_\eta}\left|\psi_1^{'}(u_{i\Delta_n}+\delta_j)-\psi_1^{'}(u_{i\Delta_n})\right|K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\left|\widetilde{X}_{(i-1)\Delta_n}-x\right|^l.$$

Under the Assumption 9 and the fact that $\left|\widetilde{X}_{(i-1)\Delta_n}-x\right| \leq Ch_n$ due to the bounded support of the kernel function, we have

$$\begin{aligned}
& E\sup_{D_\eta}|Y(\Lambda_n)| \\
\leq\ & \frac{1}{nh^{l+1}}\sum_{i=1}^n E[\sup_{D_\eta}\left|\psi_1^{'}(u_{i\Delta_n}+\delta_j)-\psi_1^{'}(u_{i\Delta_n})\right|K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\left|\widetilde{X}_{(i-1)\Delta_n}-x\right|^l] \\
=\ & \frac{1}{nh^{l+1}}\sum_{i=1}^n E\Big\{K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\left|\widetilde{X}_{(i-1)\Delta_n}-x\right|^l E[\sup_{D_\eta}\left|\psi_1^{'}(u_{i\Delta_n}+\delta_j)-\psi_1^{'}(u_{i\Delta_n})\right|X_{(i-1)\Delta_n}=x]\Big\} \\
\leq\ & o_p(1)\cdot\frac{1}{nh^{l+1}}\sum_{i=1}^n E\Big\{K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\left|\widetilde{X}_{(i-1)\Delta_n}-x\right|^l\Big\} = o_p(1),
\end{aligned}$$

Since $\max_{1\leq i\leq n}|\eta_{i\Delta_n}| = o_p(1)$, then we can obtain $Y(\hat{\Lambda}_n) = o_p(1)$ with $\hat{\Lambda}_n = (\eta_1,\eta_1,\cdots,\eta_n)^T$.

Finally, we get the result based on the above conclusions:

$$\frac{1}{nh^{l+1}}\sum_{i=1}^n \psi_1^{'}(u_{i\Delta_n}+\eta_{i\Delta_n})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l = G_1(x_0)p(x_0)K_l(1+o_p(1)),$$

that is,

$$\sum_{i=1}^n \psi_1^{'}(u_{i\Delta_n}+\eta_{i\Delta_n})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\big(\widetilde{X}_{(i-1)\Delta_n}-x\big)^l = nh^{l+1}G_1(x_0)p(x_0)K_l(1+o_p(1)).$$

The part (ii) can be obtained by the similar procedures as the detailed proof of part (i) and the Taylor's expansion, so we omit the details for simplicity ■

Lemma 4 *Under the Assumptions 1-9, we have*

$$\frac{\sqrt{\Delta_n}}{\sqrt{nh}}\begin{pmatrix}\sum_{i=1}^n \psi(u_{i\Delta_n})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big) \\ \sum_{i=1}^n \psi(u_{i\Delta_n})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\end{pmatrix} \xrightarrow{D} N(0,\Sigma_1),$$

where $\Sigma_1 = G_2(x_0)p(x_0)V$.

Proof. Denote

$$\Pi(\widetilde{X}_{i\Delta_n}) := \frac{\sqrt{\Delta_n}}{\sqrt{nh}}\begin{pmatrix}\sum_{i=1}^n \psi_1(u_{i\Delta_n})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big) \\ \sum_{i=1}^n \psi_1(u_{i\Delta_n})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)\end{pmatrix},$$

$$\Pi(X_{i\Delta_n}) := \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \begin{pmatrix} \sum_{i=1}^n \psi_1(u_{i\Delta_n}^{'}) K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big) \\ \sum_{i=1}^n \psi_1(u_{i\Delta_n}^{'}) K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big) \end{pmatrix}$$

where $u_{i\Delta_n}^{'} = \frac{X_{i\Delta_n}-X_{(i-1)\Delta_n}}{\Delta_n} - \mu(X_{(i-1)\Delta_n})$ and $X_{i\Delta_n}$ are discretely sampled from $dX_t = \mu(X_{t-})dt + \sigma(X_{t-})dW_t + \int_{\mathscr{E}} c(X_{t-}, z)r(\omega, dt, dz)$.

Write

$$\Pi(\widetilde{X}_{i\Delta_n}) = \Pi(X_{i\Delta_n}) + \big[\Pi(\widetilde{X}_{i\Delta_n}) - \Pi(X_{i\Delta_n})\big],$$

hence to prove $\Pi(\widetilde{X}_{i\Delta_n}) \xrightarrow{D} N(0, \Sigma_1)$, it is sufficient to verify that $\Pi(X_{i\Delta_n}) \xrightarrow{D} N(0, \Sigma_1)$ and $\Pi(\widetilde{X}_{i\Delta_n}) - \Pi(X_{i\Delta_n}) \xrightarrow{P} 0$.

Firstly, we prove that $\Pi(X_{i\Delta_n}) \xrightarrow{D} N(0, \Sigma_1)$.

To prove $\Pi(X_{i\Delta_n}) \xrightarrow{D} N(0, \Sigma_1)$, it is sufficient to prove that the linear combination of the two components of $\Pi(X_{i\Delta_n})$:

$$k_1\Pi(X_{i\Delta_n})_{11} + k_2\Pi(X_{i\Delta_n})_{21} := \sum_{i=1}^n q_i$$

converges in law.

We can obtain that $\sum_{i=1}^n q_i \xrightarrow{D} N(0, G_2(x)p(x)[k_1^2 J_0 + 2k_1k_2J_1 + k_2^2 J_2])$ by Lemma 2, if the following conditions hold

$|S_1| = \big|\sum_{i=1}^n E_{i-1}[q_i]\big| \xrightarrow{P} 0;$

$S_2 = \sum_{i=1}^n \big(E_{i-1}[{q_i}^2] - E_{i-1}^2[q_i]\big) \xrightarrow{P} G_2(x)p(x)[k_1^2 J_0 + 2k_1k_2J_1 + k_2^2 J_2];$

$S_3 = \sum_{i=1}^n E_{i-1}[{q_i}^4] \xrightarrow{P} 0;$

where $E_{i-1}[\ \cdot\ ] = E[\ \cdot\ |X_{t_{i-1}}]$.

We will utilize the same technical details as $\frac{1}{nh^{l+1}}T_{n1}^{'}$, hence more details are omitted for simplicity in the following proofs.

For S_1,

$$\begin{aligned}
& |S_1| \\
= \ & \Big|\frac{\sqrt{\Delta_n}}{\sqrt{nh}} \sum_{i=1}^n E_{i-1}\big[k_1\psi_1(u_{i\Delta_n}^{'})K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big) + k_2\psi_1(u_{i\Delta_n}^{'})K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\big]\Big| \\
\leq \ & \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \sum_{i=1}^n \Big|\Big\{k_1K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big) + k_2K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)E\big[\psi_1(u_{i\Delta_n}^{'})|X_{(i-1)\Delta_n}\big]\Big\}\Big| \\
= \ & \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \sum_{i=1}^n \Big|\Big[k_1K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big) + k_2K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big]O(\Delta_n)\Big| \\
= \ & \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \cdot O(\Delta_n) \sum_{i=1}^n \Big|\Big[k_1K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big) + k_2K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big]\Big| \\
= \ & O\Big(\frac{\sqrt{\Delta_n}}{\sqrt{nh}} \cdot O(\Delta_n) \cdot nh \cdot \big|k_1K_0p(x) + k_2K_1p(x)\big|\Big) \\
= \ & O\Big(\sqrt{nh\Delta_n^3}\Big) \longrightarrow 0.
\end{aligned}$$

For S_2, Similarly as S_1, $\sum_{i=1}^{n}\big(E_{i-1}^2[q_i]\big)=O(\Delta_n^3)\cdot[k_1K_0p(x)+k_2K_1p(x)]^2\longrightarrow 0$, it is sufficient to deal with the first part of S_2.

$$\begin{aligned}
& \sum_{i=1}^{n} E_{i-1}[{q_i}^2] \\
= \quad & \frac{\Delta_n}{nh}\sum_{i=1}^{n} E_{i-1}\Big\{\psi_1^2(u_{i\Delta_n}^{'})\big[k_1K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)+k_2K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\big]^2\Big\} \\
= \quad & \frac{\Delta_n}{nh}\sum_{i=1}^{n}\Big[k_1K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)+k_2K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big]^2 E_{i-1}\big[\psi_1^2(u_{i\Delta_n}^{'})\big] \\
= \quad & G_2(x)\frac{1}{nh}\sum_{i=1}^{n}\Big[k_1K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)+k_2K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big]^2 \\
\xrightarrow{P} \quad & G_2(x)p(x)[k_1^2J_0+2k_1k_2J_1+k_2^2J_2],
\end{aligned}$$

so we have proved that $S_2=\sum_{i=1}^{n}\big(E_{i-1}[{q_i}^2]-E_{i-1}^2[q_i]\big)\xrightarrow{P} G_2(x)p(x)[k_1^2J_0+2k_1k_2J_1+k_2^2J_2]$.

For S_3,

$$\begin{aligned}
& \sum_{i=1}^{n} E_{i-1}[{q_i}^4] \\
= \quad & \frac{\Delta_n^2}{(nh)^2}\sum_{i=1}^{n} E_{i-1}\Big\{\psi_1^4(u_{i\Delta_n}^{'})\big[k_1K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)+k_2K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\big]^4\Big\} \\
= \quad & \frac{1}{\Delta_n nh}\cdot\frac{1}{nh}\sum_{i=1}^{n}\Big[k_1K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)+k_2K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big)\Big]^4 \Delta_n^3E_{i-1}\big[\psi_1^4(u_{i\Delta_n}^{'})\big] \\
= \quad & \frac{1}{\Delta_n nh}\cdot O_p(1)\xrightarrow{P} 0.
\end{aligned}$$

Finally, we prove $\Pi(\widetilde{X}_{i\Delta_n})-\Pi(X_{i\Delta_n})\xrightarrow{P} 0$, for which we only consider the first component here due to the fact that we can prove the second component by $K(x)\cdot x$ instead of $K(x)$.

$$\begin{aligned}
& \Pi_{11}(\widetilde{X}_{i\Delta_n})-\Pi_{11}(X_{i\Delta_n}) \\
= \quad & \frac{\sqrt{\Delta_n}}{\sqrt{nh}}\sum_{i=1}^{n}\psi_1(u_{i\Delta_n})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)-\frac{\sqrt{\Delta_n}}{\sqrt{nh}}\sum_{i=1}^{n}\psi_1(u_{i\Delta_n}^{'})K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big) \\
= \quad & \frac{\sqrt{\Delta_n}}{\sqrt{nh}}\sum_{i=1}^{n}\psi_1(u_{i\Delta_n})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)-\frac{\sqrt{\Delta_n}}{\sqrt{nh}}\sum_{i=1}^{n}\psi_1(u_{i\Delta_n}^{'})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big) \\
+ \quad & \frac{\sqrt{\Delta_n}}{\sqrt{nh}}\sum_{i=1}^{n}\psi_1(u_{i\Delta_n}^{'})K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n}-x}{h}\Big)-\frac{\sqrt{\Delta_n}}{\sqrt{nh}}\sum_{i=1}^{n}\psi_1(u_{i\Delta_n}^{'})K\Big(\frac{X_{(i-1)\Delta_n}-x}{h}\Big) \\
:= \quad & \delta_{3n}+\delta_{4n}.
\end{aligned}$$

For δ_{3n},

$$\begin{aligned}
E[|\delta_{3n}|] &= \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \sum_{i=1}^{n} E\Big[K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)|\psi_1(u_{i\Delta_n}) - \psi_1(u'_{i\Delta_n})|\Big] \\
&= \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \sum_{i=1}^{n} E\Big\{K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)E\Big[|\psi_1(u_{i\Delta_n}) - \psi_1(u'_{i\Delta_n})||X_{(i-1)\Delta_n} = x\Big]\Big\} \\
&= O(\Delta_n) \cdot \sqrt{n\Delta_n h_n} E\Big[\frac{1}{h_n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big)\Big] \\
&= O(\sqrt{n\Delta_n^3 h_n}) \to 0,
\end{aligned}$$

because

$$E\big[u_{i\Delta_n} - u'_{i\Delta_n}\big] = E\Big[\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - \frac{X_{i\Delta_n} - X_{(i-1)\Delta_n}}{\Delta_n}\Big] = O\big(\Delta_n\big).$$

For δ_{4n}, it suffices to show that

$$\lim_{n\to\infty} E(\delta_{4n}) = 0, \qquad \lim_{n\to\infty} Var(\delta_{4n}) = 0.$$

By the mean-value theorem, we have

$$\begin{aligned}
E(\delta_{4n}) &= \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \sum_{i=1}^{n} E\Big[\psi_1(u'_{i\Delta_n})\Big(K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h}\Big) - K\Big(\frac{X_{(i-1)\Delta_n} - x}{h}\Big)\Big)\Big] \\
&= \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \sum_{i=1}^{n} E\Big[\psi_1(u'_{i\Delta_n})K'(\xi_{ni})\frac{\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n}}{h}\Big] \\
&= \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \sum_{i=1}^{n} E\Big\{K'(\xi_{ni})\frac{\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n}}{h} E[\psi_1(u'_{i\Delta_n})|X_{(i-1)\Delta_n} = x]\Big\} \\
&= O(\Delta_n) \cdot \frac{\sqrt{\Delta_n}}{\sqrt{nh}} \sum_{i=1}^{n} E\Big[K'(\xi_{ni})\frac{\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n}}{h}\Big] \\
&= O(\Delta_n) \cdot \frac{\sqrt{n\Delta_n}}{\sqrt{h}} E\Big[K'(\xi_{ni})\frac{\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n}}{h}\Big],
\end{aligned}$$

Hence, by the equation (30),

$$\begin{aligned}
|E(\delta_{4n})| &\le O(\Delta_n) \cdot \frac{\sqrt{n\Delta_n}}{\sqrt{h}} E\Big[|K'(\xi_{ni})|\frac{|\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n}|}{h}\Big] \\
&\le O(\Delta_n) \cdot \frac{\sqrt{n\Delta_n}}{\sqrt{h}} \sqrt{\Delta_n \log(1/\Delta_n)} E\Big[\frac{1}{h}|K'(\xi_{ni})|\Big] \\
&= O(\Delta_n \cdot \frac{\sqrt{n\Delta_n^2 \log(1/\Delta_n)}}{\sqrt{h}}) \to 0.
\end{aligned}$$

Moreover,

$$\begin{aligned} Var(\delta_{4n}) &= \frac{1}{h_n^2} Var\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{\Delta_n}\psi_1(u'_{i\Delta_n})\frac{1}{\sqrt{h_n}}K'(\xi_{ni})\big(\widetilde{X}_{(i-1)\Delta_n}-X_{(i-1)\Delta_n}\big)\Big] \\ &:= \frac{1}{h_n^2} Var\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i\Big] \end{aligned}$$

We find $Var[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i] = \frac{1}{n}\sum_{i=1}^{n} Var[f_i] + \varepsilon_n$ where ε_n represents the sum of $\frac{2}{n}\sum_{j=1}^{n-1}\sum_{i=j+1}^{n}$ terms involving the autovariances.

Firstly, we prove that $E[f_i^2] < \infty$. Now we calculate $E[f_i^2]$:

$$\begin{aligned} E[f_i^2] &= E\Big[\Delta_n\psi_1^2(u'_{i\Delta_n})\frac{1}{h_n}K'^2(\xi_{ni})\big(\widetilde{X}_{(i-1)\Delta_n}-X_{(i-1)\Delta_n}\big)^2\Big] \\ &= E\Big\{\frac{1}{h_n}K'^2(\xi_{ni})\big(\widetilde{X}_{(i-1)\Delta_n}-X_{(i-1)\Delta_n}\big)^2 E\Big[\Delta_n\psi_1^2(u'_{i\Delta_n})|X_{(i-1)\Delta_n}=x\Big]\Big\} \\ &= G_2(x)E\Big[\frac{1}{h_n}K'^2(\xi_{ni})\big(\widetilde{X}_{(i-1)\Delta_n}-X_{(i-1)\Delta_n}\big)^2\Big] \\ &\le G_2(x)\Delta_n\log(1/\Delta_n)E\Big[\frac{1}{h_n}K'^2(\xi_{ni})\Big] \\ &= O\big(\Delta_n\log(1/\Delta_n)\big) < \infty. \end{aligned}$$

Then, we prove that the series $\varepsilon_n = O(\frac{1}{\Delta_n^\alpha})$. It is known that $\{X_{i\Delta_n}, i = 1,2,\cdots\}$ and $\{\tilde{X}_{i\Delta_n}, i = 1,2,\cdots\}$ are stationary and $\rho-$mixing with the same size. Moreover, measurable functions of $\rho-$mixing processes are $\rho-$mixing with the same size in the space of square integrable functions according to the defination of $\rho-$mixing.

$$f_i := \sqrt{\Delta_n}\psi_1(u'_{i\Delta_n})\frac{1}{\sqrt{h_n}}K'(\xi_{ni})\big(\widetilde{X}_{(i-1)\Delta_n}-X_{(i-1)\Delta_n}\big).$$

We have already proved $E[f_i^2] < \infty$ above, so f_i is stationary under Assumption 2 and $\rho-$mixing with the same size as $\{X_{i\Delta_n}, i = 1,2,\cdots\}$ and $\{\tilde{X}_{i\Delta_n}, i = 1,2,\cdots\}$. Hence $|\varepsilon_n| \le \frac{8}{n}\sum_{j=1}^{n-1}\sum_{i=j+1}^{n}\rho((i-j)\Delta_n)Ef_2^2 = O(\frac{Ef_2^2}{\Delta_n^\alpha})$. So the series $\varepsilon_n = O(\frac{\Delta_n\log(1/\Delta_n)}{\Delta_n^\alpha})$.
One easily obtains $Var[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i] = O(\frac{\Delta_n\log(1/\Delta_n)}{\Delta_n^\alpha})$. In conclusion,

$$Var[\delta_{4n}] = \frac{1}{h_n^2}Var[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i] = O\Big(\frac{\Delta_n\log(1/\Delta_n)}{h_n^2\Delta_n^\alpha}\Big) \to 0 \ as \ n \to \infty.$$

So $\delta_{4n} \to 0$ can be deduced from the above considerations. ■

6.2 The proof of Theorem 1

Proof. (i) Firstly, we prove the consistency of the local M-estimators for $\mu(x)$ and $\mu^{'}(x)$. Write

$$L_n(r) = \sum_{i=1}^{n} \rho_1\Big(\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - a_1 - b_1\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big),$$

and

$$r = (a_1, hb_1)^T, \qquad r_0 = (\mu(x_0), h\mu^{'}(x_0))^T.$$

Denote $S_\varepsilon = \{r : ||r - r_0|| = \varepsilon\}$: the circle centered at r_0 with radius ε, we will prove that for any sufficiently small ε,

$$\lim_{n\to\infty} P\{\inf_{r\in S_\varepsilon} L_n(r) > L_n(r_0) = 0\} = 1. \tag{31}$$

We have

$$\begin{aligned}
r_{i\Delta_n} &= (r - r_0)^T \begin{pmatrix} 1 \\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} \end{pmatrix} \\
&= (a_1 - \mu(x_0), hb_1 - h\mu^{'}(x_0)) \begin{pmatrix} 1 \\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} \end{pmatrix} \\
&= a_1 - \mu(x_0) + (hb_1 - h\mu^{'}(x_0))\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} \\
&= a_1 - \mu(x_0) + (b_1 - \mu^{'}(x_0))(\widetilde{X}_{(i-1)\Delta_n} - x_0) \\
&= a_1 + b_1(\widetilde{X}_{(i-1)\Delta_n} - x_0) - \mu(x_0) - \mu^{'}(x_0)(\widetilde{X}_{(i-1)\Delta_n} - x_0) \\
&= a_1 + b_1(\widetilde{X}_{(i-1)\Delta_n} - x_0) + R_1(\widetilde{X}_{(i-1)\Delta_n}) - \mu(\widetilde{X}_{(i-1)\Delta_n}) \\
&= a_1 + b_1(\widetilde{X}_{(i-1)\Delta_n} - x_0) + R_1(\widetilde{X}_{(i-1)\Delta_n}) - \Big(\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - u_{i\Delta_n}\Big) \\
&\quad + \big[\mu(X_{(i-1)\Delta_n}) - \mu(\widetilde{X}_{(i-1)\Delta_n})\big],
\end{aligned}$$

where $R_1(\widetilde{X}_{(i-1)\Delta_n}) = \mu(\widetilde{X}_{(i-1)\Delta_n}) - \mu(x_0) - \mu^{'}(x_0)(\widetilde{X}_{(i-1)\Delta_n} - x_0)$.

For $r \in S_\varepsilon$, we have

$$
\begin{aligned}
& L_n(r) - L_n(r_0) \\
= & \sum_{i=1}^{n} \rho_1\Big(\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - a_1 - b_1\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \\
& - \sum_{i=1}^{n} \rho_1\Big(\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - \mu(x_0) - \mu^{'}(x_0)\big(\widetilde{X}_{(i-1)\Delta_n} - x\big)\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \\
= & \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big)\Big[\rho_1\Big(u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n}) - r_{i\Delta_n} + \epsilon_1\Big) - \\
& \qquad \rho_1\Big(u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n}) + \epsilon_1\Big)\Big] \\
= & \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \int_{u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n})}^{u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n}) - r_{i\Delta_n}} \psi_1(s + \epsilon_1) ds \\
= & \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \int_{u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n})}^{u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n}) - r_{i\Delta_n}} \Big\{\psi_1(u_{i\Delta_n}) + \psi_1^{'}(u_{i\Delta_n})(s - u_{i\Delta_n}) + \\
+ & \Big[\psi_1(s + \epsilon_1) - \psi_1(u_{i\Delta_n}) - \psi_1^{'}(u_{i\Delta_n})(s - u_{i\Delta_n})\Big]\Big\} ds \\
:= & L_{n1} + L_{n2} + L_{n3},
\end{aligned}
$$

where $\epsilon_1 = \mu(X_{(i-1)\Delta_n}) - \mu(\widetilde{X}_{(i-1)\Delta_n})$.
In the subsequent part, we will show three asymptotic results:

$$L_{n1} = o_p(nh\varepsilon), \tag{32}$$

$$L_{n2} = \frac{nh}{2}(r - r_0)^T G_1(x_0) p(x_0) U (1 + o_p(1))(r - r_0) + O_p(nh^3\varepsilon), \tag{33}$$

$$L_{n2} = o_p(nh\varepsilon^2). \tag{34}$$

As for L_{n1}, we have

$$
\begin{aligned}
L_{n1} &= \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \int_{u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n})}^{u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n}) - r_{i\Delta_n}} \psi_1(u_{i\Delta_n}) ds \\
&= \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \psi_1(u_{i\Delta_n})(-r_{i\Delta_n}) \\
&= -(r - r_0)^T \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \psi_1(u_{i\Delta_n}) \begin{pmatrix} 1 \\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} \end{pmatrix} \\
&= -(r - r_0)^T W_n,
\end{aligned}
$$

where

$$W_n = \begin{pmatrix} \sum_{i=1}^n \psi_1(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \\ \sum_{i=1}^n \psi_1(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} \end{pmatrix}.$$

According to Lemma 4, we have $E(W_n) = 0$ and

$$Var(W_n) = \frac{nh}{\Delta_n} G_2(x_0)p(x_0)V(1 + o_p(1)).$$

So we have $W_n = O_p(\sqrt{\frac{nh}{\Delta_n}}) = o_p(\sqrt{nh})$ by the fact of $W_n = E(W_n) + O_p\big(\sqrt{Var(W_n)}\big)$ and the assumption $nh\Delta_n \to 0$, which implies that (32) holds for $r \in S_\varepsilon$.

As for L_{n2}, we have

$$\begin{aligned}
L_{n2} &= \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \int_{u_{i\Delta_n}+R_1(\widetilde{X}_{(i-1)\Delta_n})}^{u_{i\Delta_n}+R_1(\widetilde{X}_{(i-1)\Delta_n})-r_{i\Delta_n}} \psi_1^{'}(u_{i\Delta_n})(s-u_{i\Delta_n})ds \\
&= \frac{1}{2}\sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big)\psi_1^{'}(u_{i\Delta_n})\big[\big(R_1(\widetilde{X}_{(i-1)\Delta_n}) - r_{i\Delta_n}\big)^2 - R_1^2(\widetilde{X}_{(i-1)\Delta_n})\big] \\
&= \frac{1}{2}\sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big)\psi_1^{'}(u_{i\Delta_n})\big[r_{i\Delta_n}^2 - 2R_1(\widetilde{X}_{(i-1)\Delta_n})r_{i\Delta_n}\big] \\
&= \frac{1}{2}\sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big)\psi_1^{'}(u_{i\Delta_n})(r-r_0)^T \begin{pmatrix} 1 & \frac{\widetilde{X}_{(i-1)\Delta_n}-x_0}{h_n} \\ \frac{\widetilde{X}_{(i-1)\Delta_n}-x_0}{h_n} & \frac{\left(\widetilde{X}_{(i-1)\Delta_n}-x_0\right)^2}{h_n^2} \end{pmatrix} (r-r_0) \\
&\quad -\sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big)\psi_1^{'}(u_{i\Delta_n})R_1(\widetilde{X}_{(i-1)\Delta_n})r_{i\Delta_n} \\
&:= L_{n21} + L_{n22}.
\end{aligned}$$

Applying Lemma 3 with $l = 0, 1, 2$ respectively, it can be similarly obtained

$$\begin{aligned}
L_{n21} &= \frac{1}{2}\sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big)\psi_1^{'}(u_{i\Delta_n})(r-r_0)^T \begin{pmatrix} 1 & \frac{\widetilde{X}_{(i-1)\Delta_n}-x_0}{h_n} \\ \frac{\widetilde{X}_{(i-1)\Delta_n}-x_0}{h_n} & \frac{\left(\widetilde{X}_{(i-1)\Delta_n}-x_0\right)^2}{h_n^2} \end{pmatrix} (r-r_0) \\
&= \frac{nh}{2}(r-r_0)^T G_1(x_0)p((x_0)) \begin{pmatrix} K_0 & K_1 \\ K_1 & K_2 \end{pmatrix} (1+o_p(1))(r-r_0) \\
&= \frac{nh}{2}(r-r_0)^T G_1(x_0)p(x_0)U(1+o_p(1))(r-r_0)
\end{aligned}$$

and

$$\begin{aligned}
L_{n22} &= -\sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big)\psi_1^{'}(u_{i\Delta_n})R_1(\widetilde{X}_{(i-1)\Delta_n})r_{i\Delta_n} \\
&= -(r-r_0)^T \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big)\psi_1^{'}(u_{i\Delta_n})R_1(\widetilde{X}_{(i-1)\Delta_n}) \begin{pmatrix} 1 \\ \frac{\widetilde{X}_{(i-1)\Delta_n}-x_0}{h_n} \end{pmatrix} \\
&= -\frac{nh_n^3}{2}(r-r_0)^T G_1(x_0)\mu^{''}((x_0))p(x_0) \begin{pmatrix} K_2 \\ K_3 \end{pmatrix} (1+o_p(1)) \\
&= O_p(nh^3\varepsilon).
\end{aligned}$$

As a conclusion,

$$L_{n2} = L_{n21} + L_{n22} = \frac{nh}{2}(r-r_0)^T G_1(x_0)p(x_0)U(1+o_p(1))(r-r_0) + O_p(nh^3\varepsilon).$$

As for L_{n3}, we get

$$\begin{aligned}
& L_{n3} \\
= \ & \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \int_{u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n})}^{u_{i\Delta_n} + R_1(\widetilde{X}_{(i-1)\Delta_n}) - r_{i\Delta_n}} \Big[\psi_1(s+\epsilon_1) - \psi_1(u_{i\Delta_n}) - \psi_1^{'}(u_{i\Delta_n})(s - u_{i\Delta_n})\Big] ds \\
= \ & \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \int_{R_1(\widetilde{X}_{(i-1)\Delta_n})}^{R_1(\widetilde{X}_{(i-1)\Delta_n}) - r_{i\Delta_n}} \Big[\psi_1(u_{i\Delta_n} + s + \epsilon_1) - \psi_1(u_{i\Delta_n}) - \psi_1^{'}(u_{i\Delta_n})s\Big] ds \\
= \ & \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \Big[\psi_1(u_{i\Delta_n} + z_{i\Delta_n} + \epsilon_1) - \psi_1(u_{i\Delta_n}) - \psi_1^{'}(u_{i\Delta_n}) z_{i\Delta_n}\Big](-r_{i\Delta_n}) \\
= \ & -(r-r_0)^T \sum_{i=1}^{n} K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \times \\
& \times \Big[\psi_1(u_{i\Delta_n} + z_{i\Delta_n} + \epsilon_1) - \psi_1(u_{i\Delta_n}) - \psi_1^{'}(u_{i\Delta_n}) z_{i\Delta_n}\Big] \begin{pmatrix} 1 \\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} \end{pmatrix}
\end{aligned}$$

where the second equation follows from the mean-value theorem and $z_{i\Delta_n}$ lies between $R_1(\widetilde{X}_{(i-1)\Delta_n}) - r_{i\Delta_n}$ and $R_1(\widetilde{X}_{(i-1)\Delta_n})$ for $i = 1, 2, \cdots, n$.

For $\big|\widetilde{X}_{(i-1)\Delta_n} - x\big| \leq h_n$ based on the bounded compact of $K(\cdot)$, we have

$$\begin{aligned}
\max_i \big|z_{i\Delta_n} + \epsilon_1\big| \ &\leq \ \max_i \big|R_1(\widetilde{X}_{(i-1)\Delta_n})\big| + \Big|(r-r_0)^T \begin{pmatrix} 1 \\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} \end{pmatrix}\Big| + |\epsilon_1| \\
&\leq \ \max_i \big|R_1(\widetilde{X}_{(i-1)\Delta_n})\big| + 2\varepsilon + C\sqrt{\Delta_n log(1/\Delta_n)}, \qquad (35)
\end{aligned}$$

where $\epsilon_1 = \mu(X_{(i-1)\Delta_n}) - \mu(\widetilde{X}_{(i-1)\Delta_n}) = \mu^{'}(\xi_{n,i})(X_{(i-1)\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})$, $\xi_{n,i}$ lies between $X_{(i-1)\Delta_n}$ for $i = 1, 2, \cdots, n$ and $\widetilde{X}_{(i-1)\Delta_n}$ and C denotes $\big|\max_i \mu^{'}(\xi_{n,i})\big|$.

Meanwhile, by the Taylor expansion,

$$\begin{aligned}
\max_i \big|R_1(\widetilde{X}_{(i-1)\Delta_n})\big| \ &= \ \max_i \big|\mu(\widetilde{X}_{(i-1)\Delta_n}) - \mu(x_0) - \mu^{'}(x_0)(\widetilde{X}_{(i-1)\Delta_n} - x_0)\big| \\
&= \ \max_i \big|\frac{1}{2}\mu^{''}(\theta_{n,i})(\widetilde{X}_{(i-1)\Delta_n} - x_0)\big| \\
&= \ O_p(h_n^2), \qquad (36)
\end{aligned}$$

where $\theta_{n,i}$ lies between $\widetilde{X}_{(i-1)\Delta_n}$ and x_0 for $i = 1, 2, \cdots, n$.

For fixed ε (which tends to zero) and any $\Delta_n < \varepsilon^2$, by equations (35) - (36), the Assumption 9 and the similar argument as that in Lemma 3 (i) for T_{n2}, we obtain

$$L_{n2} = o_p(nh\varepsilon^2).$$

Let λ be the smallest eigenvalue of the positive definite matrix U. Then for any $r \in S_\varepsilon$ (fixed ε but sufficiently small tending to zero), we have

$$
\begin{aligned}
L_n(r) - L_n(r_0) &= L_{n1} + L_{n2} + L_{n3} \\
&= \frac{nh}{2}(r - r_0)^T G_1(x_0) p(x_0) U(1 + o_p(1))(r - r_0) + o_p(nh\varepsilon^2) \\
&\geq \frac{nh}{2} G_1(x_0) p(x_0) \lambda \varepsilon^2 (1 + o_p(1)) + o_p(nh\varepsilon^2).
\end{aligned}
$$

So we have

$$
\lim_{n\to\infty} P\Big\{ \inf_{r\in S_\varepsilon} \big(L_n(r) - L_n(r_0)\big) \geq \frac{nh}{2} G_1(x_0) p(x_0) \lambda \varepsilon^2 (1 + o_p(1)) > 0 \Big\} = 1, \quad (37)
$$

which implies that (31) holds.

By (31), $L_n(r)$ has a local minimum in the interior of S_ε, so at the local minimum, (13) and (14) must be satisfied. Let be $(\hat{\mu}_n(x_0), h\hat{\mu}_n^{'}(x_0))^T$ the closest solutions to $r_0 = (\mu(x_0), h\mu^{'}(x_0))^T$, then

$$
\lim_{n\to\infty} P\Big\{ (\hat{\mu}_n(x_0) - \mu(x_0))^2 + h_n^2 (\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0))^2 \leq \varepsilon^2 \Big\} = 1,
$$

which implies the consistency of the local M-estimators of $\hat{\mu}_n(x_0)$ and $\hat{\mu}_n^{'}(x_0)$.
(ii) Finally, we prove the normality of the local M-estimators of $\hat{\mu}_n(x_0)$ and $\hat{\mu}_n^{'}(x_0)$ for $\mu(x)$ and $\mu^{'}(x)$. Let

$$
\hat{\eta}_{i\Delta_n} = R_1(\widetilde{X}_{(i-1)\Delta_n}) - (\hat{\mu}_n(x_0) - \mu(x_0)) - (\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0))(\widetilde{X}_{(i-1)\Delta_n} - x_0), \quad (38)
$$

we have

$$
\begin{aligned}
\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} &= u_{i\Delta_n} + \mu(X_{(i-1)\Delta_n}) \\
&= u_{i\Delta_n} + \mu(X_{(i-1)\Delta_n}) - \mu(x_0) - \mu^{'}(x_0)(\widetilde{X}_{(i-1)\Delta_n} - x_0) \\
&\quad + \mu(x_0) + \mu^{'}(x_0)(\widetilde{X}_{(i-1)\Delta_n} - x_0) \\
&= u_{i\Delta_n} + [\mu(X_{(i-1)\Delta_n}) - \mu(\widetilde{X}_{(i-1)\Delta_n})] + R_1(\widetilde{X}_{(i-1)\Delta_n}) \\
&\quad + \hat{\mu}_n(x_0) + \hat{\mu}_n^{'}(x_0)(\widetilde{X}_{(i-1)\Delta_n} - x_0) + \hat{\eta}_{i\Delta_n} - R_1(\widetilde{X}_{(i-1)\Delta_n}) \\
&= u_{i\Delta_n} + [\mu(X_{(i-1)\Delta_n}) - \mu(\widetilde{X}_{(i-1)\Delta_n})] + \hat{\mu}_n(x_0) + \hat{\mu}_n^{'}(x_0)(\widetilde{X}_{(i-1)\Delta_n} - x_0) + \hat{\eta}_{i\Delta_n} \\
&= \hat{\mu}_n(x_0) + \hat{\mu}_n^{'}(x_0)(\widetilde{X}_{(i-1)\Delta_n} - x_0) + u_{i\Delta_n} + \hat{\eta}_{i\Delta_n} + \epsilon_1.
\end{aligned}
$$

Therefore, it follows from equation (13) with $a_1 = \hat{\mu}_n(x_0)$ and $b_1 = \hat{\mu}_n^{'}(x_0)$ that

$$\begin{aligned}
\begin{pmatrix}0\\0\end{pmatrix} &= \sum_{i=1}^{n} \psi_1\big(u_{i\Delta_n} + \hat{\eta}_{i\Delta_n} + \epsilon_1\big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \begin{pmatrix}1\\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\end{pmatrix}\\
&= \sum_{i=1}^{n} \Big[\psi_1(u_{i\Delta_n}) + \psi_1^{'}(u_{i\Delta_n})\hat{\eta}_{i\Delta_n} + \big[\psi_1(u_{i\Delta_n} + \hat{\eta}_{i\Delta_n} + \epsilon_1) - \psi_1(u_{i\Delta_n}) - \psi_1^{'}(u_{i\Delta_n})\hat{\eta}_{i\Delta_n}\big]\Big] \times\\
&\qquad \times K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \begin{pmatrix}1\\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\end{pmatrix}\\
&:= V_{n1} + V_{n2} + V_{n3},
\end{aligned}$$

where $V_{n1} = W_n$.

By (38) and Lemma 3, we get

$$\begin{aligned}
V_{n2} &= \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n}) R_1(\widetilde{X}_{(i-1)\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \begin{pmatrix}1\\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\end{pmatrix}\\
&\quad - \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \times\\
&\quad \times \begin{pmatrix} (\hat{\mu}_n(x_0) - \mu(x_0)) + (\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0))(\widetilde{X}_{(i-1)\Delta_n} - x_0)\\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}[(\hat{\mu}_n(x_0) - \mu(x_0)) + (\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0))(\widetilde{X}_{(i-1)\Delta_n} - x_0)]\end{pmatrix}\\
&= \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n}) R_1(\widetilde{X}_{(i-1)\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \begin{pmatrix}1\\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\end{pmatrix}\\
&\quad - \sum_{i=1}^{n} \psi_1^{'}(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \begin{pmatrix} 1 & \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} & \frac{\left(\widetilde{X}_{(i-1)\Delta_n} - x_0\right)^2}{h_n^2}\end{pmatrix} \begin{pmatrix}\hat{\mu}_n(x_0) - \mu(x_0)\\ h_n(\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0))\end{pmatrix}\\
&= \frac{nh^3}{2} G_1(x_0)\mu^{"}(x_0)p(x_0) \begin{pmatrix}K_2\\K_3\end{pmatrix} (1 + o_p(1))\\
&\quad - nhG_1(x_0)p(x_0) \begin{pmatrix}K_0 & K_1\\ K_1 & K_2\end{pmatrix} (1 + o_p(1)) \begin{pmatrix}\hat{\mu}_n(x_0) - \mu(x_0)\\ h_n(\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0))\end{pmatrix}\\
&= \frac{nh^3 G_1(x_0)\mu^{"}(x_0)p(x_0)}{2} A(1 + o_p(1)) - nhG_1(x_0)p(x_0)U(1 + o_p(1)) \begin{pmatrix}\hat{\mu}_n(x_0) - \mu(x_0)\\ h_n(\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0))\end{pmatrix}\\
&:= V_{n21} + V_{n22}.
\end{aligned}$$

Noting that for $\left|\widetilde{X}_{(i-1)\Delta_n} - x\right| \le h_n$ based on the bounded compact of $K(\cdot)$, we have

$$\begin{aligned}
\sup_i |\hat{\eta}_{i\Delta_n}| &= \sup_i \left|R_1(\widetilde{X}_{(i-1)\Delta_n}) - (\hat{\mu}_n(x_0) - \mu(x_0)) - (\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0))(\widetilde{X}_{(i-1)\Delta_n} - x_0)\right|\\
&\le \sup_i |R_1(\widetilde{X}_{(i-1)\Delta_n})| + |\hat{\mu}_n(x_0) - \mu(x_0)| + h_n|\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)|\\
&= O_p\big(h_n^2 + |\hat{\mu}_n(x_0) - \mu(x_0)| + h_n|\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)|\big)\\
&= o_p(1),
\end{aligned}$$

where the last equality follows from the consistency of $(\hat{\mu}_n(x_0), h_n\hat{\mu}_n^{'}(x_0))$.

Then, by the Assumption 9, Assumption 6 and the similar argument as that in Lemma 3 for T_{n2} or Theorem 1 (i) for L_{n3}, we obtain

$$\begin{aligned}
& V_{n3} \\
= \ & \sum_{i=1}^{n} \big[\psi_1(u_{i\Delta_n} + \hat{\eta}_{i\Delta_n} + \epsilon_1) - \psi_1(u_{i\Delta_n}) - \psi_1^{'}(u_{i\Delta_n})\hat{\eta}_{i\Delta_n}\big] K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n}\Big) \begin{pmatrix} 1 \\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x}{h_n} \end{pmatrix} \\
= \ & o_p(nh)\big[h_n^2 + |\hat{\mu}_n(x_0) - \mu(x_0)| + h_n|\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)| + c\sqrt{\Delta_n log(1/\Delta_n)}\big] + c\cdot nh\sqrt{\Delta_n log(1/\Delta_n)} \\
= \ & o_p(nh)\big[h_n^2 + |\hat{\mu}_n(x_0) - \mu(x_0)| + h_n|\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)|\big] \\
= \ & o_p(V_{n22}).
\end{aligned}$$

Therefore, by the equation $V_{n1} + V_{n2} + V_{n3} = 0$, we have

$$\begin{pmatrix} \hat{\mu}_n(x_0) - \mu(x_0) \\ h_n(\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)) \end{pmatrix} = \frac{1}{nh} G_1^{-1}(x_0)p^{-1}(x_0)U^{-1}(1+o_p(1))W_n + \frac{h_n^2}{2}\mu^{''}(x_0)U^{-1}A(1+o_p(1)),$$

which follows that

$$\begin{aligned}
& \sqrt{nh\Delta_n}\Big[\begin{pmatrix} \hat{\mu}_n(x_0) - \mu(x_0) \\ h_n(\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)) \end{pmatrix} - \frac{h_n^2}{2}\mu^{''}(x_0)U^{-1}A(1+o_p(1))\Big] \\
= \ & G_1^{-1}(x_0)p^{-1}(x_0)U^{-1}(1+o_p(1))\frac{\sqrt{\Delta_n}}{\sqrt{nh}}W_n.
\end{aligned}$$

Finally, by Lemma 4, Assumption 6 and the Slutsky's Theorem, we can prove

$$\begin{aligned}
& \sqrt{nh\Delta_n}\Big[\begin{pmatrix} \hat{\mu}_n(x_0) - \mu(x_0) \\ h_n(\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)) \end{pmatrix} - \frac{h_n^2}{2}\mu^{''}(x_0)U^{-1}A(1+o_p(1))\Big] \\
= \ & G_1^{-1}(x_0)p^{-1}(x_0)U^{-1}(1+o_p(1))\frac{\sqrt{\Delta_n}}{\sqrt{nh}}W_n \\
\xrightarrow{D} \ & G_1^{-1}(x_0)p^{-1}(x_0)U^{-1}N(0,\Sigma_1) \\
= \ & N(0, \frac{G_2(x_0)}{G_1^2(x_0)p(x_0)}U^{-1}VU^{-1}) \\
:= \ & N(0,\Sigma_2).
\end{aligned}$$

We have proved the consistency and the normality of the local M-estimators: $\hat{\mu}_n(x_0)$ and $\hat{\mu}_n^{'}(x_0)$ for $\mu(x)$ and $\mu^{'}(x)$. ■

6.3 The proof of Theorem 2

Proof. For simplicity, denote H_n as $\begin{pmatrix} 1 & 0 \\ 0 & h_n \end{pmatrix}$.

By (15), we deduce

$$\begin{pmatrix} \widetilde{\mu}_n(x_0) - \mu(x_0) \\ h_n\big(\widetilde{\mu}_n^{'}(x_0) - \mu^{'}(x_0)\big) \end{pmatrix} = \begin{pmatrix} \hat{\mu}_0(x_0) - \mu(x_0) \\ h_n\big(\hat{\mu}_0^{'}(x_0) - \mu^{'}(x_0)\big) \end{pmatrix} - H_n W_n^{-1}\Psi_n(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0)), \tag{39}$$

where

$$W_n = \begin{pmatrix} \frac{\partial}{\partial a_1}\Psi_{n1}(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0)), & \frac{\partial}{\partial b_1}\Psi_{n1}(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0)) \\ \frac{\partial}{\partial a_1}\Psi_{n2}(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0)), & \frac{\partial}{\partial b_1}\Psi_{n2}(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0)) \end{pmatrix} := \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}.$$

Let

$$\hat{\delta}_{i\Delta_n} = R_1(\widetilde{X}_{(i-1)\Delta_n}) - (\hat{\mu}_0(x_0) - \mu(x_0)) - (\hat{\mu}_0^{'}(x_0) - \mu^{'}(x_0))(\widetilde{X}_{(i-1)\Delta_n} - x_0),$$

then by the definitions of Ψ_{n1} , Ψ_{n2} and Lemma 3 (i), we have

$$\begin{aligned} w_{11} &= -\sum_{i=1}^{n} \psi_1^{'}\Big(\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - \hat{\mu}_0(x_0) - \hat{\mu}_0^{'}(x_0)\big(\widetilde{X}_{(i-1)\Delta_n} - x_0\big)\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\Big) \\ &= -\sum_{i=1}^{n} \psi_1^{'}\Big(u_{i\Delta_n} + \hat{\delta}_{i\Delta_n} + \epsilon_1\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\Big) \\ &= -nh_n G_1(x_0)p(x_0)K_0(1 + o_p(1)), \end{aligned}$$

which follows from the fact that for $\big|\widetilde{X}_{(i-1)\Delta_n} - x\big| \leq h_n$ based on the bounded compact of $K(\cdot)$, we have

$$\begin{aligned} \sup_i |\hat{\delta}_{i\Delta_n}| &= \sup_i \big|R_1(\widetilde{X}_{(i-1)\Delta_n}) - (\hat{\mu}_0(x_0) - \mu(x_0)) - (\hat{\mu}_0^{'}(x_0) - \mu^{'}(x_0))(\widetilde{X}_{(i-1)\Delta_n} - x_0)\big| \\ &\leq \sup_i |R_1(\widetilde{X}_{(i-1)\Delta_n})| + |\hat{\mu}_n(x_0) - \mu(x_0)| + h_n|\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)| \\ &= O_p\big(h_n^2 + |\hat{\mu}_n(x_0) - \mu(x_0)| + h_n|\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)|\big) \\ &= O_p\Big(h_n^2 + \frac{1}{\sqrt{nh_n\Delta_n}}\Big) = o_p(1), \end{aligned}$$

and $|\epsilon_1| = |\mu(X_{(i-1)\Delta_n}) - \mu(\widetilde{X}_{(i-1)\Delta_n})| = |\mu^{'}(\xi_{n,i})(X_{(i-1)\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})| = C\sqrt{\Delta_n log(1/\Delta_n)} = o_p\Big(h_n^2 + \frac{1}{\sqrt{nh_n\Delta_n}}\Big)$.
In the similar arguments, we can obtain

$$w_{12} = w_{21} = -nh_n^2 G_1(x_0)p(x_0)K_1(1 + o_p(1))$$

and

$$w_{22} = -nh_n^3 G_1(x_0)p(x_0)K_2(1 + o_p(1)).$$

Combination of the above results leads to

$$W_n = -H_n \begin{pmatrix} K_0 & K_1 \\ K_1 & K_2 \end{pmatrix} H_n nh_n G_1(x_0)p(x_0)(1 + o_p(1))$$

and

$$W_n^{-1} = -H_n^{-1}(K_0K_2 - K_1^2)^{-1} \begin{pmatrix} K_2 & -K_1 \\ -K_1 & K_0 \end{pmatrix} H_n^{-1}(nh_n G_1(x_0)p(x_0))^{-1}(1 + o_p(1)).$$

Meanwhile, according to the denotation of $\hat{\delta}_{i\Delta_n}$ and the expression of Ψ_{n1}, we get

$$
\begin{aligned}
& \Psi_{n1}(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0)) \\
= \quad & \sum_{i=1}^{n} \psi_1\Big(\frac{\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n}}{\Delta_n} - \hat{\mu}_0(x_0) - \hat{\mu}_0^{'}(x_0)\big(\widetilde{X}_{(i-1)\Delta_n} - x_0\big)\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\Big) \\
= \quad & \sum_{i=1}^{n} \psi_1\Big(u_{i\Delta_n} + \hat{\delta}_{i\Delta_n} + \epsilon_1\Big) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\Big) \\
= \quad & \sum_{i=1}^{n} \Big[\psi_1(u_{i\Delta_n}) + \psi_1^{'}(u_{i\Delta_n})\hat{\delta}_{i\Delta_n} + \big[\psi_1(u_{i\Delta_n} + \hat{\delta}_{i\Delta_n} + \epsilon_1) - \psi_1(u_{i\Delta_n}) - \psi_1^{'}(u_{i\Delta_n})\hat{\delta}_{i\Delta_n}\big]\Big] \times \\
& \qquad\qquad \times K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\Big) \\
:= \quad & I_{n1} + I_{n2} + I_{n3}.
\end{aligned}
$$

By virtue of Lemma 3 (ii) with $\hat{\eta}_{i\Delta_n} = 0$, it yields

$$
\begin{aligned}
& I_{n2} \\
= \quad & -nh_n G_1(x_0)p(x_0)K_0\big(\hat{\mu}_0(x_0) - \mu(x_0)\big)(1 + o_p(1)) - nh_n^2 G_1(x_0)p(x_0)K_1\big(\hat{\mu}_0^{'}(x_0) - \mu^{'}(x_0)\big)(1 + o_p(1)) \\
& + \frac{G_1(x_0)}{2} nh_n^3 K_2 \mu^{''}(x_0)p(x_0)(1 + o_p(1)) \\
= \quad & -nh_n G_1(x_0)p(x_0)(1 + o_p(1))(K_0, K_1)\begin{pmatrix} \hat{\mu}_0(x_0) - \mu(x_0) \\ h_n\big(\hat{\mu}_0^{'}(x_0) - \mu^{'}(x_0)\big) \end{pmatrix} \\
& + \frac{G_1(x_0)}{2} nh_n^3 K_2 \mu^{''}(x_0)p(x_0)(1 + o_p(1)).
\end{aligned}
$$

Similarly as the procedure of V_{n3}, we can prove that

$$
I_{n3} = o_p(nh)\big[h_n^2 + |\hat{\mu}_n(x_0) - \mu(x_0)| + h_n|\hat{\mu}_n^{'}(x_0) - \mu^{'}(x_0)|\big].
$$

Substituting the results for I_{n1}, I_{n2}, I_{n3}, we obtain

$$
\begin{aligned}
\Psi_{n1}(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0)) \quad = \quad & \frac{G_1(x_0)}{2} nh_n^3 K_2 \mu^{''}(x_0)p(x_0)(1 + o_p(1)) + \sum_{i=1}^{n} \psi_1(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\Big) \\
& -nh_n G_1(x_0)p(x_0)(1 + o_p(1))(K_0, K_1)\begin{pmatrix} \hat{\mu}_0(x_0) - \mu(x_0) \\ h_n\big(\hat{\mu}_0^{'}(x_0) - \mu^{'}(x_0)\big) \end{pmatrix}. \qquad (40)
\end{aligned}
$$

In the similar way, we can show

$$
\begin{aligned}
& \Psi_{n2}(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0)) \\
= \quad & \frac{G_1(x_0)}{2} nh_n^4 K_3 \mu^{''}(x_0)p(x_0)(1 + o_p(1)) + \sum_{i=1}^{n} \psi_1(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\Big)\big(\widetilde{X}_{(i-1)\Delta_n} - x_0\big) \\
& -nh_n^2 G_1(x_0)p(x_0)(1 + o_p(1))(K_1, K_2)\begin{pmatrix} \hat{\mu}_0(x_0) - \mu(x_0) \\ h_n\big(\hat{\mu}_0^{'}(x_0) - \mu^{'}(x_0)\big) \end{pmatrix}. \qquad (41)
\end{aligned}
$$

Substituting $\Psi_{n1}(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0))$ and $\Psi_{n2}(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0))$ into $H_n W_n^{-1}\Psi_n(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0))$, it follows

$$\begin{aligned}
H_n W_n^{-1}\Psi_n(\hat{\mu}_0(x_0), \hat{\mu}_0^{'}(x_0)) &= H_n W_n^{-1}\frac{G_1(x_0)}{2} n h_n^3 \mu^{''}(x_0) p(x_0)(1+o_p(1)) H_n \begin{pmatrix} K_2 \\ K_3 \end{pmatrix} \\
&\quad - H_n W_n^{-1} n h_n G_1(x_0) p(x_0) H_n \begin{pmatrix} K_0 & K_1 \\ K_1 & K_2 \end{pmatrix} \begin{pmatrix} \hat{\mu}_0(x_0) - \mu(x_0) \\ h_n\big(\hat{\mu}_0^{'}(x_0) - \mu^{'}(x_0)\big) \end{pmatrix} \\
&\quad + H_n W_n^{-1} H_n \sum_{i=1}^{n} \psi_1(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\Big) \begin{pmatrix} 1 \\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} \end{pmatrix} \\
&:= J_{n1} + J_{n2} + J_{n3}. \qquad (42)
\end{aligned}$$

Under a simple calculation, we can achieve

$$J_{n1} = -\frac{\mu^{''}(x_0) h_n^2}{2(K_0K_2 - K_1^2)} \begin{pmatrix} K_2^2 - K_1K_3 \\ K_0K_3 - K_1K_2 \end{pmatrix} (1+o_p(1))$$

and

$$J_{n2} = \begin{pmatrix} \hat{\mu}_0(x_0) - \mu(x_0) \\ h_n\big(\hat{\mu}_0^{'}(x_0) - \mu^{'}(x_0)\big) \end{pmatrix} (1+o_p(1)).$$

Hence, by (39) and (42), we get

$$\begin{aligned}
&\begin{pmatrix} \widetilde{\mu}_n(x_0) - \mu(x_0) \\ h(\widetilde{\mu}_n^{'}(x_0) - \mu^{'}(x_0)) \end{pmatrix} \\
= \;& (n h_n G_1(x_0) p(x_0))^{-1} \begin{pmatrix} K_0 & K_1 \\ K_1 & K_2 \end{pmatrix}^{-1} \sum_{i=1}^{n} \psi_1(u_{i\Delta_n}) K\Big(\frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n}\Big) \begin{pmatrix} 1 \\ \frac{\widetilde{X}_{(i-1)\Delta_n} - x_0}{h_n} \end{pmatrix} \\
& + \frac{\mu^{''}(x_0) h_n^2}{2(K_0K_2 - K_1^2)} \begin{pmatrix} K_2^2 - K_1K_3 \\ K_0K_3 - K_1K_2 \end{pmatrix} + o_p\Big(h_n^2 + \frac{1}{\sqrt{n h_n \Delta_n}}\Big). \qquad (43)
\end{aligned}$$

The conclusion

$$\sqrt{nh\Delta_n}\Big[\begin{pmatrix} \widetilde{\mu}_n(x_0) - \mu(x_0) \\ h(\widetilde{\mu}_n^{'}(x_0) - \mu^{'}(x_0)) \end{pmatrix} - \frac{h^2\mu^{''}(x_0)}{2} U^{-1} A\Big] \xrightarrow{D} N(0, \Sigma_2)$$

follows from (43) and Lemma 4, where $\Sigma_2 = \frac{G_2(x_0)}{G_1^2(x_0)p(x_0)} U^{-1} V U^{-1}$. ■

References

[1] Aït-Sahalia, Y. and Jacod, J. Testing for jumps in a discretely observed process. *Annals of Statistics*, (2009), 37, 184-222.

[2] Bandi, F. and Nguyen, T. On the functional estimation of jump-diffusion models. *Journal of Econometrics* , (2003), 116, 293-328.

[3] Bandi, F. and Phillips, P. Fully nonparametric estimation of scalar diffusion models. *Econometrica*, (2003), 71, 241-283.

[4] Barndorff-Nielsen, O.E. and Shephard, N. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, (2006), 4, 1-30.

[5] Bickel, P. One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.*, (1975), 70, 428-434.

[6] Cont, R. and Tankov, P. Financial modeling with jump processes. Chapman and Hall/CRC (2004).

[7] Chen, Y. and Zhang, L. Local linear estimation of second-order jump-diffusion model. *Communications in Statistics - Theory and Methods*, (2015), 44, 3903-3920.

[8] Fan, J. and Zhang, C. A reexamination of diffusion estimators with applications to financial model validation. *Journal of the American Statistical Association*, (2003), 98, 118-134.

[9] Florens-Zmirou, D. On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.* , (1993), 30, 790-804.

[10] Fan, J. and Jiang, J. Variable bandwidth and one-step local M-estimator. *Science in China (Series A)* , (2000), 43, 65-81.

[11] Gugushvili, S. and Spereij, P. Parametric inference for stochastic differential equations: a smooth and match approach. *Latin American Journal of Probability & Mathematical Statistics* , (2012), 9, 609-635.

[12] Hall, P. and Jones, M. Adaptive M-estimation in nonparametric regression. *Ann. Statist.* , (1990), 18, 1712-1728.

[13] Huber, P. Robust Regression. *Ann. Statist.* , (1973), 1, 799-821.

[14] Jacod, J. *Statistics and high-frequency data.* In: Statistical Methods for Stochastic Differential Equations, CRC Boca Raton-London-New York, 191-310, (2012).

[15] Jacod, J. and Shiryaev, A. *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* **288**. Springer, Berlin, (2003).

[16] Jiang, M. and Mack, Y. Robust local polynomial regression for dependent data. *Sinica*, (2001), 11, 705-722.

[17] Johannes, M.S. The economic and statistical role of jumps to interest rates. *Journal of Finance*, (2004), 59, 227-260.

[18] Kessler, M. Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, (1997), 24, 211-229.

[19] Protter, P. *Stochastic integration and differential equations*, 2nd ed. *Applications of Mathematics (New York)* **21**. Springer, Berlin, (2004).

[20] Nicolau, J. Nonparametric estimation of second-order stochastic differential equations. *Econometric theory*, (2007), 23, 880-898.

[21] Park, J. and Phillips, P. Nonlinear regressions with integrated time series. *Econometrica*, (2001), 69, 117-161.

[22] Rogers, L. and Williams, D. Diffusions, Markov processes and Martingales: Volume 2, *Itô calculus*, Cambridge University Press (2000).

[23] Shimizu, Y. and Yoshida, N. Estimation of parameters for diffusion processes with jumps from discrete observations. *Stat. Inference Stoch. Process.*, (2006), 9, 227-277.

[24] Song, Y. and Lin, Z. Empirical likelihood inference for the second-order jump-diffusion model. *Statist. Probab. Lett.*, (2013), 83, 184-195.

[25] Song, Y., Lin, Z. and Wang, H. Re-weighted Nadaraya-Watson estimation of second-order jump-diffusion model *J. Statist. Plann. Inference*, (2013), 143, 730-744. .

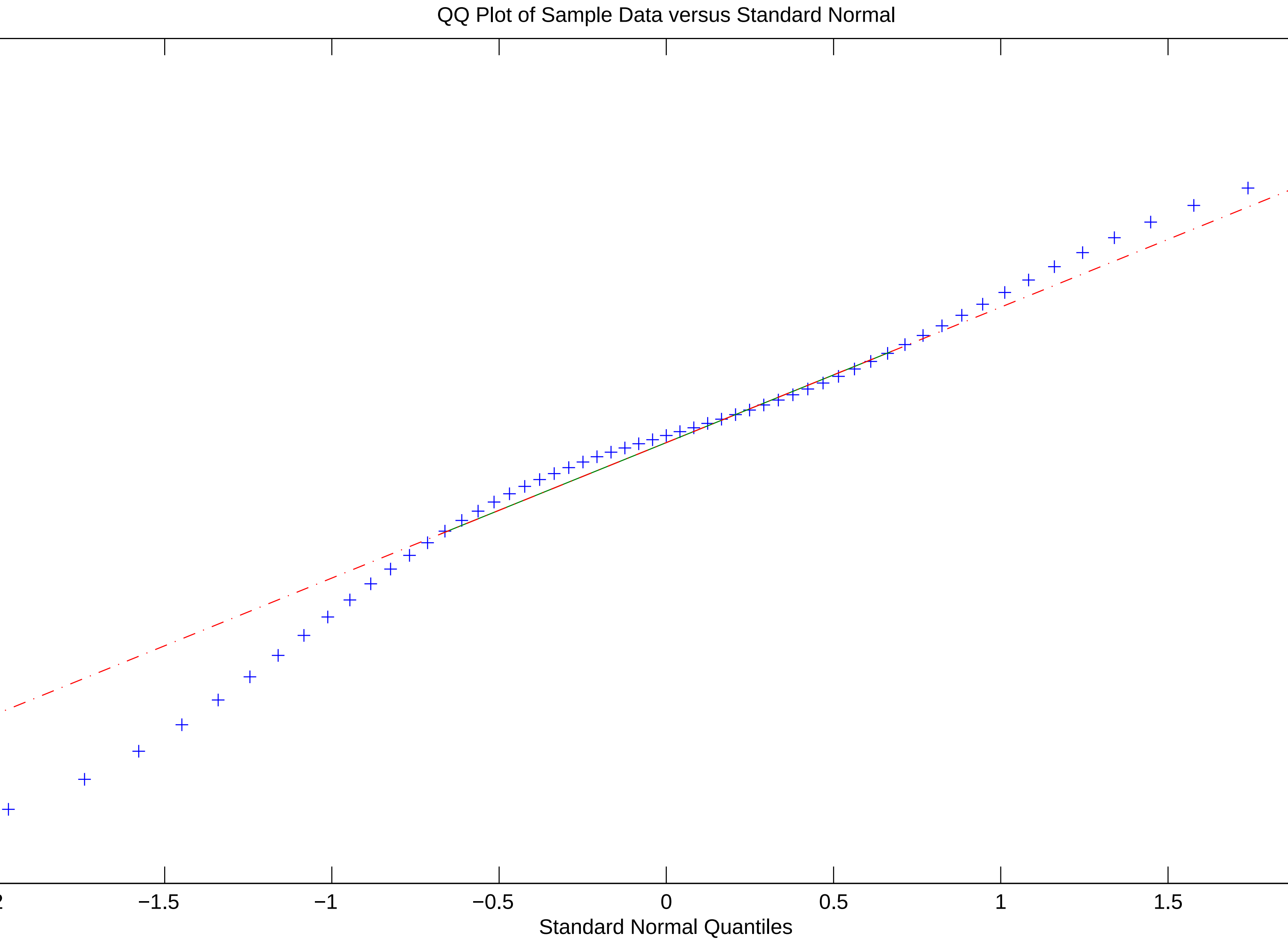
[26] Wang, H. and Lin, Z. Local linear estimation of second-order diffusion models *Comm. Statist. Theory Methods*, (2011), 40, 394-407.

[27] Wang, Y., Zhang, L. and Tang, M. Local M-estimation for jump-diffusion processes. *Statistics and Probability Letters*, (2012), 82, 1273-1284.

[28] Wang, Y., Zhang, L. and Tang, M. Re-weighted functional estimation of second-order diffusion processes. *Metrika*, (2012), 75, 1129-1151.

[29] Wang, Y. and Tang, M. Nonparametric robust function estimation for integrated diffusion processes. *Journal of Nonlinear Science and Application*, (2016), 9, 3048-3065.

[30] Xu, K. and Phillips, P. Tilted nonparametric estimation of volatility functions with empirical aplications. *ournal of Business* & *Economic Statistics*, (2011),29, 518-528.


QQ Plot of Sample Data versus Standard Normal
−1.5
−1
−0.5
0
0.5
1
1.5
Standard Normal Quantiles